



University  
of Glasgow

<https://theses.gla.ac.uk/>

Theses Digitisation:

<https://www.gla.ac.uk/myglasgow/research/enlighten/theses/digitisation/>

This is a digitised version of the original print thesis.

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses

<https://theses.gla.ac.uk/>  
[research-enlighten@glasgow.ac.uk](mailto:research-enlighten@glasgow.ac.uk)

An Investigation of the Molecular Genetics of  
Oxytetracycline Biosynthesis and Resistance in  
*Streptomyces rimosus*.

A thesis submitted for the degree of  
Doctor of Philosophy at the University of Glasgow

by

Kenneth John McDowall

Institute of Genetics  
The University  
Church Street  
Glasgow, G11 5JS

January 1991

ProQuest Number: 10983753

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10983753

Published by ProQuest LLC (2018). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code  
Microform Edition © ProQuest LLC.

ProQuest LLC.  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106 – 1346

The research reported in this thesis is my own original work except where otherwise stated and has not been submitted for any other degree.

Kenneth J McDowall, January 1991.



**Dedicated to mum and dad.**

CONTENTS	(i)
ABBREVIATIONS	(vi)
ACKNOWLEDGEMENTS	(viii)
SUMMARY	(ix)

## CONTENTS

### CHAPTER 1 INTRODUCTION

1.1 Introduction to <i>Streptomyces</i>	1
1.2 <i>Streptomyces</i> Life Cycle	3
1.3 Developmental Biology of <i>Streptomyces coelicolor</i>	5
1.4 Medical and Commercial Interest in <i>Streptomyces</i>	8
1.5 Organisation of the Genes for Antibiotic Biosynthesis	9
1.6 Pathway-specific Regulatory Genes	12
1.7 Catabolite Regulation of Secondary Metabolism	14
1.8 Role of "Signal" Molecules in Controlling Development	21
1.9 RNA Polymerase Heterogeneity in <i>Streptomyces</i>	23
1.10 Polyketide Biosynthesis	26
1.11 Actinorhodin Biosynthesis	32
1.12 Organisation of Polyketide Synthase Genes	34
1.13 Module Hypothesis	36
1.14 Biosynthesis of Oxytetracycline	38
1.15 Scope of Present work	46

### CHAPTER 2 MATERIALS AND METHODS

2.1 Bacterial Strains and Vectors	47
2.2 Microbiological Techniques and Standard Media	50
2.2.1 Standard Media for the Propagation of <i>E.coli</i>	50
2.2.2 Standard Media for the Propagation of <i>Streptomyces</i>	50

2.2.3	Sterilization	52
2.2.4	Preparation of <i>Streptomyces</i> Spore Suspensions	52
2.2.5	<i>E.coli</i> Growth Conditions	53
2.2.6	Growth of <i>Streptomyces</i> Mycelium	53
2.2.7	Quantification of <i>Streptomyces</i> Growth	53
2.2.8	Quantification of Oxytetracycline in Culture Supernatants	54
2.2.9	Antibiotics and Indicators	54
2.2.10	Preservation of <i>Streptomyces</i> and <i>E. coli</i> Strains	55
2.2.11	Introduction of Plasmid DNA into <i>E.coli</i>	55
2.2.12	Transfection of <i>E. coli</i> TG1 with Bacteriophage M13	56
2.2.13	Introduction of Plasmid DNA into <i>Streptomyces</i>	57
2.3	Nucleic Acid Isolation and Manipulation	59
2.3.1	Isolation of Plasmid DNA from <i>E.coli</i> and <i>Streptomyces</i>	59
2.3.2	Isolation of Total RNA	61
2.3.3	Quantification of Nucleic Acids	63
2.3.4	Precipitation of Nucleic Acid using Ethanol or Isopropanol	64
2.3.5	Digestion of DNA with Restriction Endonucleases	64
2.3.6	Ligation of DNA Fragments	64
2.3.7	Removal of the 5' Phosphate from Linearised DNA	64
2.3.8	Removal of Protein from Nucleic Acid Solutions using Organic Solvents	65
2.3.9	Preparation of Radio-Labelled Probes	65
2.3.9.1	Random Primed DNA Labelling Method	65
2.3.9.2	5'-End Labelling of Oligonucleotides	66
2.3.9.3	Sephadex G50 Column Chromatograph	66
2.3.10	Agarose Gel Electrophoresis	66

2.3.10.1 Mini Gels	67
2.2.10.2 Large Gels	67
2.3.10.3 Photograph of Resolved Nucleic Acid	68
2.3.11 Denaturing Polyacrylamide Gel Electrophoresis	68
2.3.12 Recovery of DNA from Agarose Gels	69
2.3.12.1 GeneClean™	69
2.3.12.2 Elecro-Elution	70
2.3.13 Isolation of DNA from Polyacrylamide Gels	70
2.3.14 <i>E.coli</i> Colony Transfer to Amersham Hybond-N™ Membrane	71
2.3.15 Southern Analysis	72
2.3.15.1 Capillary Transfer	72
2.3.15.2 Aqueous Hybridisation Conditions	72
2.3.15.2a Using Random Primed Probes	72
2.3.15.2b Using Oligonucleotides	73
2.3.16 Preparation of Single-Stranded M13 DNA	74
2.3.17 DNA Sequencing	75
2.3.18 Transcript Mapping using Complementary Single-Stranded DNA Probes	76
2.3.19 Primer Extension Mapping of Transcript 5' Ends	79
2.3.20 Autoradiography	80
2.3.21 Densitrometry Scanning	80

## CHAPTER 3 ANALYSIS OF PROMOTER ACTIVITY UPSTREAM OF *OTRA*

3.1 Introduction	81
3.1.1 Objectives	82
3.1.2 Methology	83
3.2 Results	
3.2.1 Construction of Recombinants for <i>in vivo</i> Promoter Analysis	85
3.2.2 Assay of Promoter Activity from pIJ486 Constructs	88

3.2.3 S1 Nuclease Mapping of the Putative <i>otrA</i> p2 Transcript	89
3.2.4 Transcription of <i>otrA</i> during OTC Production	93
3.3 Discussion	102
 CHAPTER 4 THE <i>OTCZ</i> GENE: NUCLEOTIDE SEQUENCE, DEDUCED FUNCTION AND TRANSCRIPTION	
4.1 Introduction	113
4.2 Results	
4.2.1 Sequencing Method	115
4.2.2 Sequencing Strategy	118
4.2.3 Computer-Assisted Sequence Analysis	125
4.2.3.1 Potential Protein Coding Regions	126
4.2.3.2 Amino Acid Sequence Comparisons	131
4.2.4 Analysis of Transcription of <i>otcZ</i>	136
4.3 Discussion	139
 CHAPTER 5 CHARACTERISATION OF THE <i>OTCC/OTCX</i> PROMOTER REGION	
5.1 Introduction	142
5.1.1 Overall Strategy	145
5.2 Results and Discussion	146
5.2.1 Construction of a Promoter Library in pIJ2843	146
5.2.2 Assay of <i>In Vivo</i> Promoter Activity	150
5.2.3 Characterisation of the <i>xylE</i> <sup>+</sup> Transformants	150
5.2.4 Location and Characterisation of the Promoter Region Responsible for the Transcription of <i>otrA</i> During Antibiotic Production	155
5.2.4.1 Location of the 5' Termini of the Divergent Transcripts	157
5.2.4.2 Quantification of the 5' Termini of the Divergent Transcripts	163

5.2.5 Comparison of the <i>otcCp1/otcXp1</i> Promoter Sequences	166
CHAPTER 6 THE <i>SPHI</i> <sub>12</sub> - <i>SSTI</i> <sub>13</sub> REGION: NUCLEOTIDE SEQUENCE AND LOCATION OF DNA ELEMENTS DIRECTING THE EXPRESSION OF <i>XYLE</i> IN pIJ2843 CONSTRUCTS	
6.1 Introduction	174
6.2 Results	177
6.2.1 DNA Sequencing of the <i>SphI</i> - <i>SstI</i> Region	177
6.2.2 Features of the Nucleotide Sequence	179
6.2.3 Analysis of the Deduced Amino Acid Sequence of PPCR1 and PPCR2	185
6.2.4 Location of the DNA Elements from the <i>SphI</i> - <i>SstI</i> Region which Direct Transcription of <i>xyle</i> in pIJ2843 Constructs	185
6.2.5 Analysis of Transcription within the <i>Sau3AI</i> <sub>c-e</sub> and <i>Sau3AI</i> <sub>f-g</sub> Regions	189
6.3 Discussion	189
CHAPTER 7 TRANSLATION INITIATION	
7.1 Introduction	196
7.2 Results and Discussion	198
7.2.1 Comparison of 16S rRNA Sequences from <i>Streptomyces</i> and <i>E. coli</i>	198
7.2.2 Complementarity Between Downstream Sequences and the 5' End of 16S rRNA	199
7.2.3 Nucleotides of 16S rRNA Implicated in Binding to Downstream Regions	205
7.3 Scope of Future Studies	206
CHAPTER 8 CONCLUDING REMARKS	209
BIBLIOGRAPHY	215

## ABBREVIATIONS

ap	ampicillin
APS	ammonium persulphate
ATP	adenosine 5'-triphosphate
bp	base pairs
Ci	curie
CIAP	calf intestinal alkaline phosphatase
cpm	counts per minute
D	dalton
DEPC	diethyl pyrocarbonate
DMSO	dimethylsulphoxide
DMF	dimethylformamide
DTT	dithiothreitol
dNTP	deoxynucleotide 5'-triphosphate
dATP	deoxyadenosine 5'-triphosphate
dCTP	deoxycytidine 5'-triphosphate
dGTP	deoxyguanosine 5'-triphosphate
dTTP	deoxythymidine 5'-triphosphate
DNA	deoxyribonucleic acid
ssDNA	single stranded deoxyribonucleic acid
dH <sub>2</sub> O	distilled water
E	core RNA polymerase
EDTA	ethylene diaminetetra-acetic acid (disodium salt)
EtBr	ethidium bromide
EtOH	ethanol
g	centrifugal force equivalent to gravitational acceleration
IPTG	isopropylthio- $\beta$ -D-galactoside
hrs	hours
kb	kilo bases
k <sub>m</sub>	kanamycin
l	litres
M	molar
mins	minutes

Mr	molecular weight
OD	optical density
ORF	open reading frame
OTC	oxytetracycline
°C	degrees centigrade
nt	nucleotide
PEG	polyethylene glycol
P <sub>i</sub>	orthophosphate
PIPES	piperazine-NN'-bis-2-ethane sulphonic acid
PPCR	potential protein coding region
RBS	ribosome binding site
RNA	ribonucleic acid
mRNA	messenger ribonucleic acid
rRNA	ribosomal ribonucleic acid
tRNA	transfer ribonucleic acid
rpm	revolutions per minute
RT	room temperature
sp. act.	specific activity
SDS	sodium dodecylsulphate
TC	tetracycline
TEMED	NNN-N'-tetramethylethylenediamine
TES	N-tris (hydroxymethyl) methyl-2-aminoethanesulphonic acid
Tris	tris (hydroxymethyl) methylamine
TSB	tryptone soya broth
tsr	thiostrepton
U	units
UV	ultraviolet
V	volts
v/v	volume by volume
w/v	weight by volume
W	watts
Xgal	5-bromo-4-chloro-3-indoyl-β-galactoside
YEME	Yeast extract-malt extract



## ACKNOWLEDGEMENTS

I express first my gratitude to Iain Hunter for his friendship, support and words of wisdom over the last three years.

Special thanks is also extended to "Auntie" Marion for keeping the laboratory running smoothly and providing me with blankets when I arrived in Glasgow to keep me warm during the cold winter nights. Needless to say the media ladies deserve a mention for preparing many batches of foul smelling media, while always remaining cheerful. Thank you Margaret and Linda for deciphering my writing and typing the reference section.

I also acknowledge my colleagues in the *Streptomyces* group for their friendship and invaluable discussion. Alison is due a special mention for proof-reading early drafts of this manuscript and reminding me of a few rules of my own language.

Thanks also to Mike Butler for arranging my visit to the Pfizer Laboratories, where I undertook several experiments described in this thesis. I am also grateful for the help and advice given by Craig Binnie and Mel Warren at Sandwich.

Last, but certainly not least, I thank my wife, Janine, for her love and understanding during this work, particularly whilst this thesis was being written.

This work was supported by a grant from the Medical Research Council.

## SUMMARY

Prior to the work described in this thesis, an oxytetracycline (OTC) resistance gene, *otrA*, had been cloned from *S. rimosus* and sequenced. S1 nuclease mapping experiments suggested that the *otrA* gene was transcribed from two promoters, 129bp (*otrAp1*) and 339-344bp (*otrAp2*) upstream of the translation start codon.

Further transcriptional analyses described in this thesis confirmed the presence of the *otrAp1* promoter, but showed that a technical artifact had led to the erroneous assignment of *otrAp2*. The transcription of the *otrA* resistance gene was examined at different stages in the fermentation of OTC. During the period of intense vegetative growth which precedes the production of OTC, the *otrA* resistance gene was transcribed predominantly from *otrAp1*. As the growth rate slowed and the onset of antibiotic production approached, the resistance gene was transcribed increasingly as part of an *otcC-otcZ-otrA* polycistronic message, while transcription from *otrAp1* decreased. The inclusion of the *otrA* resistance gene in a transcription unit along with the biosynthetic genes *otcZ* and *otcC* probably ensures that the level of resistance increases in line with the biosynthesis of oxytetracycline. Furthermore, the delay in the production of OTC can be attributed, at least in part, to the temporal regulation of transcription.

The 1.3kb *SmaI-BcII* fragment containing the *otcZ* biosynthetic gene was sequenced. The amino acid sequence of the predicted gene product showed homology to a methyltransferase from bovine pineal glands and to a region of the tetracenomycin cluster, which is also implicated in methylation. The *otcZ* gene product is thought to catalyse the C-6 methylation of the tetracyclic nucleus to form 6-methylpretetramid.

Sequencing of a 0.96kb *SphI-SstI* fragment which was located immediately upstream of *otcC* identified two PPCRs, *otcX* PPCR1 and

PPCR2, that were encoded on the opposite strand to *otcC*, *otcZ* and *otrA*. No homology was found between the amino acid sequences of the predicted gene products and entries in the NBRF-protein and EMBL databases. The *SphI*-*SstI* fragment had been shown previously to hybridize to a region of the actinorhodin cluster of *S. coelicolor*, which was implicated in hydroxylation. It is possible that the polypeptides encoded by *otcX* PPCR1 and PPCR2 have roles in the C-4 hydroxylation of 6-methylpretetramid.

Primer extension analyses identified the 5' ends of both the mRNA of *otcX* PPCR1 and PPCR2 and the *otcC-otcZ-otrA* message. These divergent transcripts were non-overlapping. Alignment of the DNA sequences immediately upstream of the 5' ends of these transcripts revealed several conserved features, which may be involved in the initiation and regulation of transcription. The most striking of these was a tandem repeat, which had sequence similarity to the recognition sequence of the PhoB regulatory protein. Under conditions of low phosphate, PhoB activates transcription of the phosphate operon in *E. coli*. As oxytetracycline is only produced under conditions of low phosphate, it is possible that a functional analogue of PhoB regulates transcription of the *otc* cluster in *S. rimosus*.

An analysis of sequences downstream of the putative translation start codons of the genes described in this thesis and a selection from other *Streptomyces* spp. has identified at least three consecutive nucleotides in the first 10 codons of every gene, which are complementary to some part of the 5' end of 16S rRNA from *Streptomyces*. It is proposed that base pairing between the 5' end of 16S rRNA and complementary nucleotides at the 5' end of a gene maybe important in determining the efficiency of translation.

**CHAPTER 1**  
**GENERAL INTRODUCTION**

## General Introduction

The primary objective of this chapter is to provide the background to the experimental investigations, which are described in this thesis. Firstly, a review of the mechanisms, which control the global changes in gene expression that underlie the development of *Streptomyces*, will be presented. Particular emphasis will be given to the integration of secondary metabolism (generally referred to as antibiotic production) into the life cycle.

During the course of this work, conserved sequences were identified within promoters of the oxytetracycline cluster. These DNA determinants were similar to the *pho* boxes, which are found in *E. coli* promoters that are regulated by phosphate. Although a molecular genetic approach was used to analyse the production of oxytetracycline in this work, physiological factors that influence the onset of secondary metabolism and morphological differentiation, particularly phosphate limitation, will be surveyed briefly to provide the necessary background for discussion on the possible role of the DNA determinants revealed by molecular analysis.

Within this thesis, experimental evidence is presented, which suggests that the delay in the production of oxytetracycline can be attributed, at least in part, to the temporal regulation of transcription. Therefore, recent publications, which have contributed significantly to our understanding of *Streptomyces* gene expression will also be discussed.

Finally, recent advances in our understanding of the biosynthesis of oxytetracycline and related polyketide antibiotics will be reviewed.

### 1.1 Introduction to *Streptomyces*

Members of the genus *Streptomyces* are aerobic, Gram-positive soil bacteria that undergo complex morphological and physiological differentiation. *Streptomyces* genomes have been estimated by renaturation kinetics to be in the range 6 to 9 megabases (Antonov *et al.*, 1978), which is up to twice the values for *E. coli* K12 and *Bacillus*

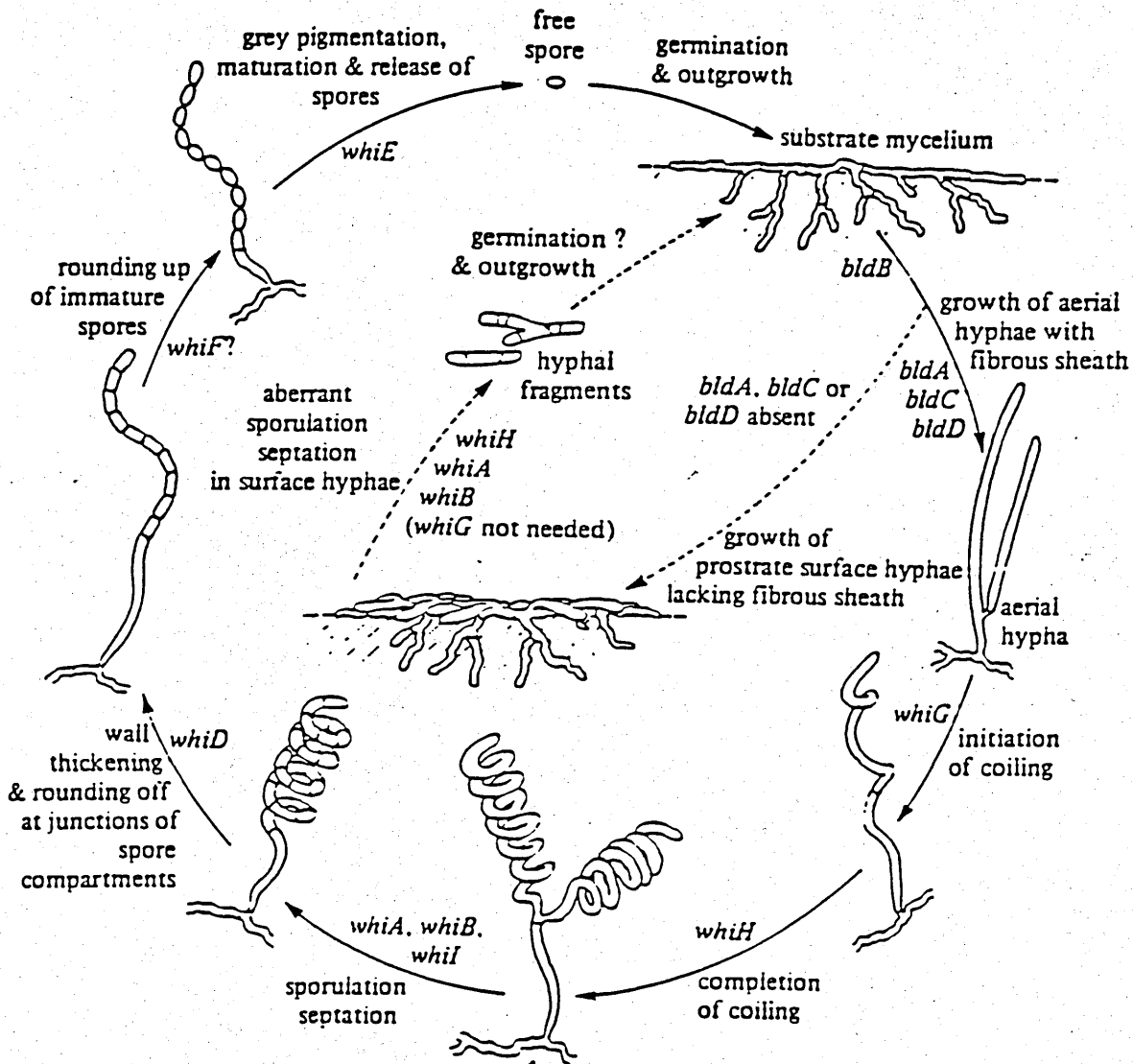


Figure 1.1. Development of *Streptomyces coelicolor*. Italicised gene symbols indicate the involvement of particular gene products in the accomplishment of given stages of development. The inner cycle (.....) occurs only with *bldA* and *bldD* mutants grown on medium containing glucose as the carbon source. During growth on mannitol phenotypic suppression of *bldA* and *bldD* mutants occurs and the outer cycle occurs. From Chater and Merrick, 1979.

*subtilis* (Hutter and Eckhardt, 1988). Recent analysis of the *S. coelicolor* chromosome, by pulse-field gel electrophoresis, does not contradict the size estimates based on kinetic analysis (cited in Hopwood and Kieser, 1990). A distinct feature of streptomycetes is the high proportion of guanine (G) and cytosine (C) in their genomes, averaging at 73 mol% G+C (Enquist and Bradley, 1971). The analysis of codon usage in *Streptomyces* shows strong bias in favour of codons with G or C in the third position (Bibb *et al.*, 1984). From a file containing 63 *Streptomyces* genes (compiled by Mervyn Bibb, unpublished work), it was shown that G residues (42%) are favoured over C residues (27%) in the first codon position, C residues (29%) are preferred over G residues (16%) in the second position and C residues (55%) are preferred to C residues (36%) in the third codon position. Therefore, the average mol G+C composition in the first, second and third position of a codon is 69%, 45% and 91% respectively.

*Streptomyces* are of interest for two principal reasons. Firstly, their morphological differentiation from vegetative mycelium to chains of spores is a relatively simple system for studying the mechanisms, which regulate the expression of developmental genes. Secondly, this genus produces a plethora of complex chemical structures, many of which have important pharmaceutical properties. These compounds have been designated "secondary metabolites" as they are not involved in the vegetative growth of the microorganism and they are thought to be produced generally as an adjunct to the differentiation process.

## 1.2 *Streptomyces* life cycle.

The *Streptomyces* life cycle (Figure 1.1), can be considered to start with the germination of a spore (Chater, 1984). The germling grows apically and branches to form a mat of irregular septate filaments, known as the substrate mycelium. When nutrients become limiting, aerial mycelium are produced from the substrate mycelium. The tips of these mycelium coil and the hyphae septate to form spores. After maturation, which involves the thickening and rounding of the spore wall, the propagules are released.

In soil and leaf litter, *Streptomyces* spp. are ecologically-important microorganisms for recycling dead plant material, which is composed largely of complex polysaccharides, lignin and cellulose. Similar to other soil microorganisms such as fungi and bacilli, the streptomycetes secrete a battery of extracellular enzymes to degrade these recalcitrant substrates.

The mycelial growth habit is well adapted to utilise these complex, insoluble polymers, which have to be degraded extracellularly before they can be taken up by the organism and used for growth. Significant factors limiting growth are, presumably, the rate of diffusion of enzymes away from the colony and the rate of diffusion of soluble degraded products back to the colony.

Mycelial outgrowth from an initial point of colonisation, as a dispersed mat probably minimises rapid nutrient exhaustion by ensuring that the growing parts of the colony are continually encountering new sources of nutrients. However, growth in a localised area does not favour dispersal of the species. Having exhausted a nutrient source, a species must be able to colonise new areas in order to have any chance of survival. In *Streptomyces*, this requirement is met by the formation of spores, which can be dispersed by air or water.

Historically, it was thought that antibiotics might play a regulatory role in differentiation. Recent evidence suggests that pamamycin stimulates the formation of aerial mycelium and may act as a signal for colonial development in the producing strain *S. alboniger* (Kondo *et al.*, 1988). This is unlikely to be a general role of secondary metabolites, however, as mutants of many *Streptomyces* spp., which have lost the ability to produce antibiotics, are viable and can complete their life cycle under laboratory conditions. By a similar argument, secondary metabolites are unlikely to be shunt products, which alleviate the potential toxic accumulation of metabolic intermediates, as growth slows and cultures enter stationary phase (as proposed by Hunter and Baumberg, 1989).



It is generally envisaged that the degradation of the substrate mycelium provides precursors and energy for the growth of the aerial mycelium. Antibiotic production is usually coincidental with the formation of the aerial mycelium. It has been argued eloquently, on the basis of teleological considerations, that the general role of secondary metabolites is to prevent invasion of the lysing substrate mycelium by motile, unicellular microorganisms, for example, bacilli (Chater, 1984). The timing of antibiotic production is clearly central to this hypothesis.

### 1.3 Developmental Biology of *Streptomyces coelicolor*

There has been a great deal of interest in the morphological development of *Streptomyces* and many mutants, blocked at different stages in the life cycle of *S. coelicolor* A3(2), have been isolated (Chater and Merrick, 1979). Two classes of mutants were identified; bald (*bld*) mutants, which do not form aerial mycelium or produce, in general, any of the antibiotics associated with this strain, and white (*whi*) mutants, which form aerial mycelium and produce antibiotics, but do not form pigmented, mature spores (Figure 1.1). With molecular genetic approaches now possible, the characterisation of these mutants is revealing interesting facets about the processes that are instrumental in *Streptomyces* development.

*BldA* mutants are unable to form aerial mycelium and produce antibiotics (Piret and Chater, 1985). Sequencing of a DNA segment which complements this mutation did not identify any potential protein coding regions (PPCR). Instead, it was deduced that this region of DNA could encode a RNA product with the structure of a leucine tRNA<sup>TTA</sup> (Lawlor *et al.*, 1987). The existing evidence suggests that TTA codons may be absent from the structural genes for morphological differentiation and antibiotic production, but present in a small subset of genes associated with the regulation of these processes. Recently, in *S. coelicolor*, it has been shown that the expression of five genes containing TTA codons is dependent on the *bldA* gene product (Leskiw *et al.*, cited in Guthrie and

Chater, 1990). This dependency provided compelling evidence that *bldA* specifies a leucine tRNA which is actively involved in protein synthesis. The rarely used TTA codon has been identified in the *actII* gene, which regulates positively actinorhodin production (Section 1.6), and in a putative actinorhodin resistance gene (Hopwood pers. comm.). Very recently, the accumulation of an undescribed transcript encoding at least one gene from the undecylprodigiosin (*red*) cluster of *S. coelicolor* A3(2) was shown to be dependent on the expression of *bldA* (Guthrie and Chater, 1990). On media containing low levels of phosphate, however, the *bldA* mutant was able to produce undecylprodigiosin. These experiments revealed that the inability of the *bldA* mutant to produce this antibiotic cannot be accounted for by the presence of TTA codons in the *red* structural genes, but rather, *bldA* influences the transcription of *red* genes. Under conditions of low phosphate, an alternative regulatory pathway (independent of *bldA*) can lead to *red* gene expression (Guthrie and Chater, 1990).

Sequencing of the DNA that complements the *whiG* mutant, revealed an ORF with the potential to encode a polypeptide with striking sequence similarity to the  $\sigma^{28}$  polypeptide from *B. subtilis* (Chater *et al.*, 1989). This suggested strongly that the promoters for a subset of genes, whose products were required for development of spores from aerial hyphae, were recognised by an alternative sigma factor,  $\sigma^{whiG}$ . In *B. subtilis* and *E. coli*, the  $\sigma^{28}$ -RNA polymerase holoenzyme transcribes genes associated with chemotaxis and motility (Chamberlain, pers. comm.), at the end of exponential growth when filamentous, multinucleate, non-motile cells divide to form uninucleate, motile cells. Mutants of *B. subtilis* deficient in  $\sigma^{28}$  remain filamentous and non-motile (Helman *et al.*, 1988). It has been suggested that, apart from motility, the above phenotype is very similar to that of *whiG* mutants, which are unable to develop beyond the extension of aerial hyphae (Chater, 1989). This family of  $\sigma$  factors may have a general role in the dispersal of bacteria during nutrient limitation.

DNA complementing the *whiE* mutant, which is deficient in spore pigmentation, has been sequenced and shown to contain ORFs that

correspond clearly to those of known polyketide synthase (PKS) genes (Section 1.12) (Davis and Chater, 1990). The biosynthesis of the grey-green spore pigment appears to be a specific feature of aerial hyphae (Davis and Chater, 1990), unlike other antibiotics which are presumed to be synthesized by the substrate mycelium. The biosynthesis of spore pigment in *S. coelicolor* may be under novel temporal and spatial control. It would be very interesting, therefore, to identify the factors necessary for the transcription of the *whiE* locus.

A model for the osmoregulation of differentiation has been proposed on the basis of the observed oscillations in the intracellular levels of glycogen (Chater, 1989). During the differentiation process, there is a cycle between the synthesis and degradation of carbon storage compounds. Glycogen granules are accumulated prior to the formation of aerial mycelium and then degraded during the extension of the aerial hyphae (Widermuth, 1970). Glycogen is again synthesized, immediately prior to the formation of the uninucleate spores (Hardisson *et al.*, 1986). The switch between synthesis and degradation is presumed to cause fluctuations in the osmotic potential of the mycelium. It is proposed that increased turgor pressure due to the degradation of glycogen granules provides the "driving force" for the extension of the aerial hyphae and the "rounding off" of the spore compartments. Similarly, the accumulation of glycogen at the end of hyphal extension is predicted to cause a reduction in turgor pressure, which permits the formation of septa. At present, it is open to debate whether the fluctuations in the abundance of glycogen has a real role in the osmoregulation of differentiation or is merely a consequence of changes in the supply and demand for cellular precursors. This model identifies the importance of the overall physiology of the mycelium and in particular, the osmotic balance. It was also proposed that the products of some *bld* genes may regulate polymer synthesis and degradation. Depending on the growth substrate, turgor pressure may be generated by metabolic pathways not regulated by *bld* genes and this could provide a possible explanation for the dependence of some *bld* phenotypes on the type of growth medium (Chater, 1989).

#### 1.4 Medical and Commercial Interests in *Streptomyces*

*Streptomyces* produce a multitude of secondary metabolites, which range from simple amino acid analogues to complex macrocyclic compounds (Goodfellow and O'Donnel, 1989). Of the six thousand or so bioactive compounds now characterised, 60% of these are produced by *Streptomyces* (Omura, 1986). However, only about 70 of the compounds derived from *Streptomyces* satisfy the high standards required clinically for efficacy and tolerability.

Many of the antibiotics produced by *Streptomyces* are of medical importance. Examples include; the  $\beta$ -lactams penicillin G and cephamycin, the tetracycline series of polyketide antibiotics, the aminoglycosides neomycin, spectinomycin and kanamycin and the peptide antibiotics actinomycin and vancomycin. Compounds produced by the streptomycetes are not only used as antibacterial agents in clinical infections. The polyethers monensin and lasalocid produced by *S. cinnamonensis* and *S. lasalocid*, respectively, are the most widely used anticoccidial agents for controlling parasitic protozoa in poultry. These compounds are used also as growth promotants in ruminants (Schumard and Callender, 1968; Mitrovic and Schildnecht, 1974). The antihelminthics avermectin and milbemycin produced by *S. avermitilis* and *Streptomyces* spp. NRRL 5370 respectively, have potent activity against nematode infections of farm animals (Burg *et al.*, 1979; Pat. No. DT2329485). The tripeptide antibiotic bialaphos, produced by *S. hygroscopicus*, has a potent herbicidal activity when cleaved by a peptidase in plants (Bayer *et al.*, 1972). The *Streptomyces* produce also a vast range of other pharmacologically-active compounds, including anti-inflammatory, vasodilatory and anti-cancer agents (Ninomija *et al.*, 1980; Blum and Carter, 1974; Yoshida *et al.*, 1982). The potential to discover secondary metabolites with novel pharmaceutical properties is an obvious reason for continuing the study of *Streptomyces*.

Recently, the ability of *Streptomyces* spp. to secrete a variety of extracellular enzymes has been the focus of considerable attention. The analysis of several genes encoding secreted polypeptides suggests

that, as in *E. coli*, a signal sequence at the amino terminus of the gene product contains the targeting information for secretion of the polypeptide. The promoter and signal sequence of a protein protease inhibitor (LTI) in *S. lividans* has been shown to direct expression and export of sCD4, a soluble derivative of the human T-cell receptor CD4 that acts as a receptor for the HIV virus (Brawner *et al.*, 1990). The vast amount of knowledge that has been gained from the development of advanced fermentation systems for antibiotic production, could benefit the development of *Streptomyces* spp. as hosts for the production and secretion of small genetically-engineered polypeptides.

### 1.5 Organisation of Genes for Antibiotic Biosynthesis

The structural genes for the biosynthesis of all the antibiotics (used here as a general term for secondary metabolites), so far studied, are closely linked in single clusters (Hunter and Baumberg, 1989). Orthogonal-field gel electrophoresis has revealed that the genes for methylenomycin production are contained on the "giant linear plasmid", SCP1 (ca. 350-520 kb) (Kinashi *et al.*, 1987) confirming the early genetic evidence (Kirby and Hopwood, 1977), which suggested that the genes had an extrachromosomal location. Apart from the gene cluster for methylenomycin A, all the other antibiotic clusters would appear to be located on the chromosome.

There have been few reports on the overall organisation of transcription within antibiotic clusters. This may reflect the difficulty experienced by many groups in identifying *Streptomyces* transcripts using Northern analysis. The overall transcriptional organisation of the actinorhodin and methylenomycin cluster has been deduced, however, by virtue of the mutational cloning strategy used to clone and analyse genes involved in antibiotic biosynthesis. Using this approach, DNA fragments inserted into an *att*-deleted derivative of actinophage  $\phi$ C31 direct integration of the vector into the homologous region of the host genome, thereby potentially disrupting gene expression. It was predicted that integration would be mutagenic when the insert was completely internal to a transcription unit (Chater and Bruton, 1983).

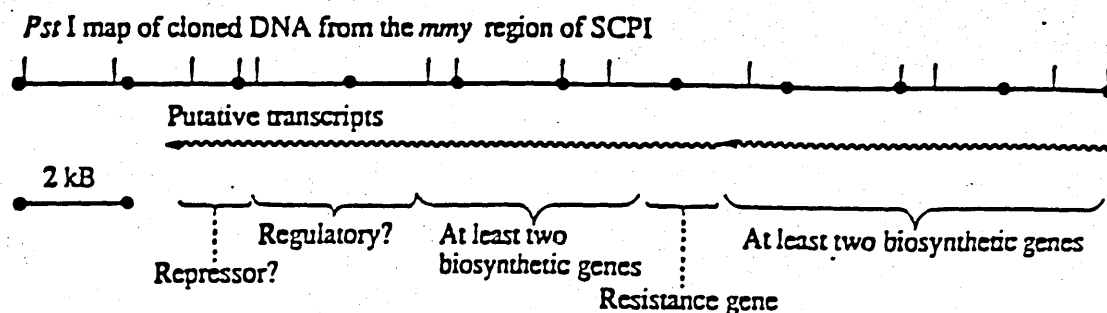


Figure 1.2. Preliminary transcription map of the methylenomycin production gene cluster. *Pst*I sites are shown by a vertical bar. From Chater and Bruton, 1985.

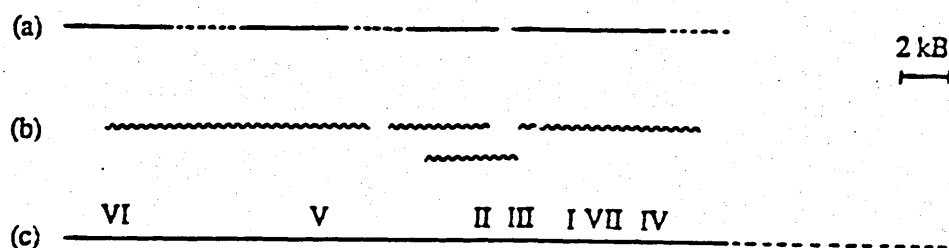


Figure 1.3. Preliminary transcription map of the gene cluster for actinorhodin production. (a) Putative transcripts from the mutational cloning experiments. (b) Putative transcripts deduced from S1 nuclease mapping experiments. (c) Location of genes as deduced by complementation tests (Malpartida and Hopwood, 1986). Modified from Hopwood *et al.*, 1986.

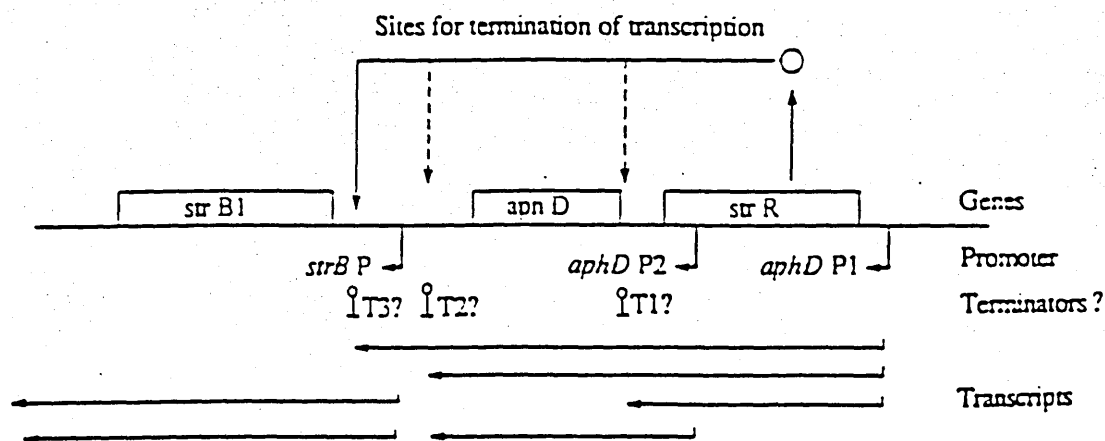


Figure 1.4. Preliminary transcription map around the *aphD* gene, which is part of the gene cluster for streptomycin production. From Distler *et al.*, 1988.

The analysis of  $\phi$ C31 constructs, which caused gene disruption, suggested that the genes for the biosynthesis of methylemomycin (the *mmy* genes) were organised into two transcription units, one of which also encoded the methylenomycin resistance (*mmr*) gene (Figure 1.2) (Chater and Bruton, 1985). Sequencing of a DNA segment encoding the *mmr* gene revealed, however, that this particular gene has to be transcribed in the opposite direction to that predicted by mutational cloning (Neal and Chater, 1987). It is not clear why an anomalous result with the mutational cloning was obtained, although the artifactual production of *mmy*-antisense RNA might be a plausible explanation (Hopwood *et al.*, 1986). The methylenomycin cluster is organised, therefore, into a least three transcription units; the central *mmr* gene being transcribed from the opposite strand to that for the longer mRNAs of the biosynthetic genes.

Mutational cloning, using *att*-deleted derivatives of  $\phi$ C31 (as described by Chater and Bruton, 1983), and S1 nuclease mapping experiments revealed, that at least four transcription units were present in the actinorhodin cluster (Figure 1.3) (Hopwood *et al.*, 1986). At least two of the transcripts (for mutant classes I,VII,IV and VI,V) would appear to be polycistronic, while another of less than 1 kb encode the *actIII* region. A fourth transcript was predicted for the *actII* region encoding the putative positive-regulatory gene (Section 1.6).

The large polycistronic messages in the actinorhodin cluster would appear to correspond to the subclustering of genes for "early" and "late" steps in the biosynthetic pathway. This feature has also been reported for the tylosin cluster (Seno and Baltz, 1988) and tetracenomycin (Motamedi and Hutchinson, 1987). These observations prompted the suggestion that subclustering may facilitate the ordered expression of genes for sequential stages in the production of the antibiotic (Motamedi and Hutchinson, 1987).

The transcriptional organisation of the streptomycin cluster around the *aphD* resistance gene has been deduced by Northern analysis (Distler *et al.*, 1987; 1988). The current working model (Figure

1.4) includes three promoters for the *strR*, *aphP* and *strB1* genes. At least seven different RNA species have been detected in this region. An anti-termination mechanism has been proposed to account for the differences in the 3' termini of transcripts originating from the same promoter (Section 1.6) (Distler *et al.*, 1988).

### 1.6 Pathway-specific Regulatory Genes

In addition to the global regulation of *Streptomyces* development (Sections 1.3 and 1.8), the biosynthesis of a number of antibiotics has been shown to be controlled by pathway-specific regulatory elements, most of which remain to be characterised satisfactorily.

Genetic analysis of the actinorhodin pathway provided the first example of a positive-acting regulatory element. The *actII* mutants were unable to produce actinorhodin in cosynthesis tests with any other classes of mutant and were presumed, therefore, to be either polar mutants within a large operon or deficient for a positive regulator (Rudd and Hopwood, 1979). The over-production of actinorhodin by the wild-type strain when DNA complementing the *actII* mutants was introduced on a plasmid, gave credence to the latter hypothesis (Hopwood, Malpartida and Chater, cited in Hopwood *et al.*, 1986). Similarly, increased production of undecylprodigiosin (red) was achieved by introducing random fragments of the *red* cluster into *S. coelicolor* on a high-copy-number plasmid (Guthrie, 1989). Interestingly, the same recombinant plasmid, containing the putative positive-regulatory (*redD*) gene, was also able to restore undecylprodigiosin production in a *bldA* mutant, providing further support for the proposal that the structural genes for the antibiotics produced by *S. coelicolor* do not contain any TTA codon (Section 1.3).

DNA segments have been isolated from *S. hygroscopicus* and *S. violaceoruber*, which complement *actII* mutants of *S. coelicolor* and are closely linked to the gene cluster for milbemycin and granaticin. These DNA segments are thought to encode positive-regulatory elements (Malpartida *et al.*, 1987). Recombinants of *S. lividans* containing the milbemycin element, *milbII*, over-produced actinorhodin (Humphreys *et*



*al.*, 1987). These observations suggest that the *actII*, *milbII* and granaticin positive-regulatory elements belong to family of regulators which have similar functions and recognise similar targets. The bialaphos cluster of *S. hygroscopicus* has also been shown to contain a positive regulator, *brpA*, which activates transcription of the biosynthetic genes (Anzai *et al.*, 1987).

A possible negative effector was identified in the methylenomycin cluster (Chater and Bruton, 1985), as mutational cloning experiments (Section 1.5), identified a region at the "leftmost" end of the cluster which caused over-production of methylenomycin when disrupted by the integration of  $\phi$ C31. The simplest explanation was that a negatively-acting regulator was encoded by the disrupted locus. The region immediately to the right of this segment also has a regulatory role since a slight increase in production resulted from gene disruption in this region (Chater and Bruton, cited in Hopwood *et al.*, 1986).

Evidence has been obtained for both positive and negative regulation of streptomycin biosynthesis in *S. griseus* (Distler *et al.*, 1988). The predicted product of the *strR* regulatory gene, which is required for the maximum expression of the *aphD* resistance gene and the *strB* amidinotransferase gene, has significant homology to the lambda Q protein involved in the antitermination of transcription. It has been proposed that the *strR* gene product facilitates transcription through a number of termination sites, which have been identified by Northern analysis (Section 1.5) (Distler *et al.*, 1988). A DNA segment has also been identified in the streptomycin cluster between the *aphD* resistance gene and the *strB* aminotransferase gene, which controls negatively expression of the *aphD* gene (Tohyama *et al.*, 1986).

In summary, the analysis of a number of different gene clusters has shown that resistance and regulatory elements are closely linked to the structural genes for the biosynthesis of antibiotics. It has also been observed that genes involved at similar stages in the biosynthesis of some antibiotics may be subclustered within the same transcription unit. Clustering probably reflects the need to coordinate the

expression of several related functions. At present, little is known about the promoters and RNA polymerase holoenzymes which are responsible for the transcription of antibiotic clusters. Although regulatory elements have been identified, the regulatory circuits of which they are a part have not yet been elucidated.

### 1.7 Catabolite Regulation of Secondary Metabolism

The biosynthesis of tetracyclines (including chloro- and oxytetracycline) and other antibiotics, including aminoglycosides, macrolides, polyenes, anthracyclines, ansamycins and polyethers is particularly sensitive to phosphate regulation (Martin *et al.*, 1988; Martin and Demain, 1980). Tetracycline can only be produced abundantly in medium containing levels of phosphate lower than that required for maximal growth of the culture. Phosphate starvation is used to limit the growth of tetracycline producers, including the Pfizer strains M4018 and M15883, and induce the production of antibiotic. The amount of phosphate required for optimum production of tetracycline is dependent on the individual production strain (Behal, 1986; 1987). The activity of anhydrotetracycline (ATC) oxygenase (encoded by *otcC* in *S. rimosus*) is reduced in cultures of *S. aureofaciens* and *S. rimosus*, when grown on medium containing levels of phosphate supraoptimal for tetracycline production (Behal, 1979; Butler pers. comm.). Although it was stated in a recent review (Liras *et al.*, 1990), no direct evidence has been published which supports the view that the repression of ATC oxygenase activity by phosphate is due to a reduction in synthesis of the enzyme. From the published data, it could be argued that phosphate regulates the catalytic efficiency of ATC oxygenase.

The synthesis of several enzymes associated with secondary metabolism (ESM) is regulated by phosphate, for example; p-aminobenzoate (PABA) synthase from the candidicin producer *S. griseus* and dTDP-glucose-4,6-dehydratase, dTDP-mycarose synthetase and macrocin O-methyltransferase from the tylosin producer *Streptomyces* T59-235 (cited in review by Liras *et al.*, 1990). In addition, the catalytic activity of some other ESM can be inhibited by growth on medium containing high levels of phosphate, for example; deacetoxycephalosporin

C synthetase in *Streptomyces clavuligerus*, *Nocardia lactamdurans* and *Acremonium chrysogenum*. This indicates that there are two levels at which phosphate control can be exerted (Liras *et al.*, 1990).

The biosynthesis of candicidin by *S. griseus* has been studied intensively as a model of phosphate control (Liras *et al.*, 1977; Martin and Demain 1977). PABA synthase activity can only be detected in *S. griseus* during the antibiotic production phase and is absent in a series of non-producing mutants (Gil *et al.*, 1985). Furthermore, this particular enzyme activity is not found during growth on medium containing high levels of phosphate. The structural gene (*pabS*) and promoter for PABA synthase have been cloned and characterised (Gil and Hopwood, 1983; Rebello *et al.*, 1989). Recently, the phosphate control of PABA synthase activity has been shown to be exerted at the level of transcription (Asturios *et al.*, 1990). The promoter for *pabS* has been reduced to a 35bp core sequence which is still sensitive to phosphate control. A hexamer, centred 10 bp upstream from the transcription start point of the *pabS* promoter, has good homology with the -10 region from the consensus sequence for the major class of eubacterial promoters. However, no homology to a canonical -35 region could be detected in the *pabS* promoter. Computer analysis of the *pabS* promoter revealed the presence of a hyphenated direct repeat on the sense strand, which overlaps with the tentative -10 region. This hyphenated repeat has significant sequence similarity to the consensus for *pho* boxes in *E. coli* (Liras *et al.*, 1990). The *pho* boxes are the DNA binding sites of the PhoB protein, which activates transcription of the phosphate regulon during phosphate limitation (Makino *et al.*, 1986, 1988). The primary *pho* boxes, in all the *E. coli* promoters regulated by phosphate, are located 10 bp upstream of the -10 regions on the same strand. It has been proposed that PhoB protein, bound to these *pho* boxes, substitutes for a typical -35 region in the phosphate-controlled promoters of *E. coli* (Makino *et al.*, 1986, 1988).

The PhoB protein belongs to a family of bacterial regulatory proteins that are an integral part of the sensory and regulatory mechanisms responsible for cellular responses to environmental changes

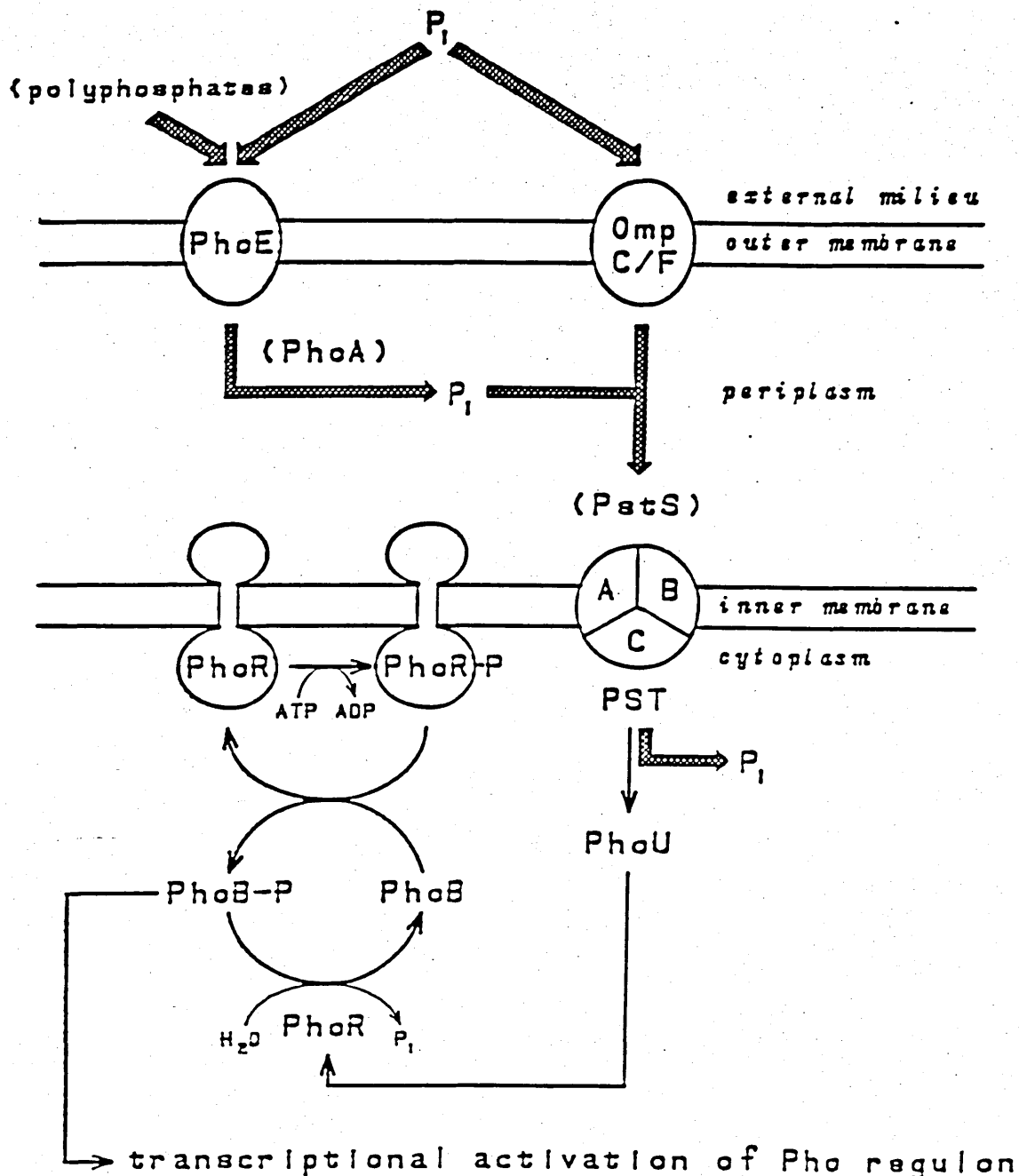


Figure 1.5. The proposed system that regulates the uptake of phosphate in *E. coli*. Extracellular phosphate ( $P_i$ ) crosses the outer membrane through the PhoE, OmpF and OmpC porins. Alternatively, phosphoesters in the periplasm are hydrolysed to  $P_i$  by alkaline phosphatase, PhoA. Under conditions of low phosphate, uptake into the cytoplasm is via the phosphate-binding protein (PstS) through the phosphate-specific transport (PST) system. The PST system in conjunction with PhoU regulates the activity of the transcriptional activator PhoB, which regulates the expression of all the operons in the Pho regulon, including *phoE*, *phoA*, *pstSCABphoU* and *phoBR*. The active form of PhoB is P-PhoB, produced by the action of the PhoR kinase. The PhoU protein together with PhoR is proposed to dephosphorylate P-PhoB. From Stock et al., 1989.

(Ronson *et al.*, 1987; Yamada *et al.*, 1989). The following summary of phosphate uptake illustrates the salient points of these regulatory mechanisms. Most of the material for this brief discussion was taken from an in depth review by Stock *et al.* (1989).

*E. coli* has a very complex mechanism for regulating the expression of genes involved in the uptake and metabolism of phosphate (Figure 1.5). In addition to the regulatory protein PhoB, another central component of the system is a membrane receptor PhoR, which phosphorylates PhoB under conditions of low phosphate, thus allowing PhoB-P to activate transcription of the phosphate operon. Frequently membrane receptor proteins are kinases that respond to environmental signals and phosphorylate response regulators that control transcription.

Active transport of phosphate across the membrane is mediated by the phosphate-specific transport (PST) system and several sugar phosphate transport systems, including glycerol-3-phosphate (encoded by the *ugp* operon) and hexose phosphates (encoded by the *uhp* operon). An anion-specific porin, PhoE, facilitates the passage of phosphate across the outer membrane into the periplasmic space. Within the periplasmic space, the PhoA alkaline phosphatase (which hydrolyses polyphosphates) and binding proteins such as PstS and UgpB, assist the delivery of phosphate to the transport systems within the cytoplasmic membrane.

Extracellular phosphate levels regulate the expression of many of the genes involved in phosphate uptake. High levels of phosphate shut down expression of the *phoA* and *phoE* genes, as well as the *pst* and *ugp* operons, while sugar phosphates such as glucose-6-phosphate and hexose phosphates induce genes that encode their own respective transport systems. Interestingly, a number of different protein kinases and response regulators are involved in controlling the expression of these genes. The PhoR protein kinase and the PhoB regulator modulate the expression of *phoA*, *phoE*, the *pst* operon and the *phoBR* operon itself. This group of genes have been designated the Pho regulon, since, all of their promoters require phosphorylated PhoB for maximum

transcription. At present, it is not known precisely how low-phosphate concentrations activate PhoR. PhoR activity does not appear to be regulated directly by either phosphate in the periplasmic space or indirectly by an interaction with PstS. An intact PTS system is essential, however, for sensing high levels of phosphate and reducing the expression of the Pho regulon.

A component of the *pst* operon PhoU, which is not involved directly in phosphate uptake, is required to shut down expression of the Pho regulon. PhoU appears to promote the dephosphorylation of PhoB-P (the active form) in response to a signal that reflects the activity of the PST. PhoM, another typical membrane receptor-protein kinase, can also phosphorylate PhoB. The PhoM kinase appears to be part of a distinct regulatory system that has some cross-specificity for PhoB.

In conclusion, the PhoR/PST system exerts both positive and negative control on PhoB. Under conditions of low phosphate, PhoR and PhoM probably both function to activate Pho expression by catalysing the transfer of phosphoryl groups to PhoB (the *pst* operon does not appear to be involved in this process). Under conditions of high phosphate, PhoR along with the PST system act to shut down the expression of the Pho regulon; in particular PhoU, which is a component of the PST system, probably activates a PhoR-dependent phosphatase that dephosphorylates PhoB-P.

Analysis of the promoter sequences of the *aphD* gene, encoding a streptomycin phosphotransferase in the streptomycin gene cluster of *S. griseus* and an open reading frame, ORFX, which is divergent to the thiostrepton-inducible *tipA* gene in *S. lividans*, has revealed tentative *pho*-like sequences in both of these phosphate-regulated promoters (Liras *et al.*, 1990). These sequence similarities prompted the suggestion that *pho*-like boxes in *Streptomyces* (also called phosphate-control (PC) sequences) may affect the binding kinetics of the RNA polymerase, possibly through a specific interaction with a putative protein bound to sequences within the vicinity of phosphate-controlled promoters.

Further to this, it was proposed that the intracellular effector which mediates phosphate control may be a highly phosphorylated nucleotide or a sugar-phosphate that acts as a sensor of the phosphate levels in the culture medium (Liras *et al.*, 1990).

Pairs of protein kinases and response regulators are involved in numerous, dynamic responses to changing environmental conditions, for example; chemotaxis and repellent stimuli in enterobacteria (Hess *et al.*, 1988), regulation of outer membrane porin expression in response to osmolarity and other conditions in *E. coli* (Igo *et al.*, 1989a; 1989b) and regulation of gene expression in response to nitrogen deprivation (Nifa *et al.*, 1988). Sporulation in *B. subtilis* also involves response regulators Spo0A and Spo0F, which are homologous to PhoB (Smith, 1989). Mutations in *spo0A* and *spo0F* have been identified which prevent initiation of the sporulation process at the onset of stationary phase when the growth of the cells becomes nutrient-limited. It is possible, therefore, that some of the *bld* mutants of *S. coelicolor* (described in Section 1.3) could be deficient in proteins related to the protein kinases and response regulators described above.

From *in vitro* assays as well as genetic analysis, it would appear that many protein kinases can phosphorylate response regulators from different systems. For example, PhoB can be phosphorylated by both PhoR and PhoM. The gene encoding PhoM is not part of the Pho operon. It has been suggested by Stock and other workers in this field that "cross-talk" between parallel phosphotransfer pathways may provide a global regulatory network for sensing the overall physiological state of the organism (Stock *et al.*, 1989). It is conceivable that integrated phosphotransferase systems may be important in controlling the switch to morphological differentiation and secondary metabolism in *Streptomyces*. Should this be the case, it could provide an additional explanation for the growth-medium dependence of some *bld* phenotypes (Section 1.3) (Chater, 1989).

The "physiological" state of *Streptomyces* is presumed to be the key factor determining the switch to antibiotic production and

morphological differentiation. It is proposed that some *bld* mutations may disrupt phosphotransfer pathways, which sense the level of nutrients. Assuming that the phosphotransferase systems for different nutrients are integrated and have a cumulative effect on the decision to undergo differentiation, the result of a mutation, which disrupts a single phosphotransferase system, may depend on the growth medium. For example, a hypothetical mutant, which disrupts results in the organism always "believing" that it is supplied well with phosphate, may undergo the complete life cycle when it is grown on medium containing high levels of phosphate. The depletion of readily utilisable carbon and/or nitrogen source may be sufficient to "trigger" differentiation. It is possible, however, that the same mutant might not undergo differentiation when grown on a medium containing good carbon and nitrogen sources but, low levels of phosphate, as the organism does not "sense" the phosphate limitation, which would "trigger" differentiation in the wild-type strain. The above example is only intended to illustrate the possible role of putative sensing mechanisms in determining the switch to differentiation. It should be noted that there is no evidence to support this model.

Physiological analysis of antibiotic production has revealed that secondary metabolism can be regulated by the availability of carbon and nitrogen, as well as phosphate (Martin and Demain, 1980; Campbell, 1986). The expression of phenoxazinone synthase (PHS) of *S. antibioticus* that forms the phenoxazinone nucleus of actinomycin is regulated by glucose (Jones, 1984). Transcriptional analysis revealed that the increase in the specific activity of PHS, observed in medium with galactose as a carbon source was due, at least in part, to increased production of PHS mRNA. The repression of PHS activity caused by glucose was also shown to be exerted at the level of transcription (Jones, 1984). Interestingly, high levels of glucose do not repress the biosynthesis of PHS in *S. lividans*. It has also been suggested that glucose represses puromycin biosynthesis in *Streptomyces alboniger* (Sankaran *et al.*, 1975) and represses transiently erythromycin production in *Saccharopolyspora erythraea* (Escalante *et al.*, 1982). In the latter example, the repression of synthesis of ESM rather than the inhibition of the catalytic activities



of some or all of the ESM was inferred, as glucose added after the onset of antibiotic production had no effect. The repression of synthesis of deacetoxycephalosporin C synthase has been proposed to be the key factor responsible for the suppression of cephamycin biosynthesis by glucose (Cortes *et al.*, 1986). Glucose is not the only carbon source which can have a repressive effect. For example, glycerol can repress cephalosporin production by *S. clavuligerus* (Aharonwitz and Demain, 1978) and citrate can repress novobiocin production by *S. niveus* (Kominek, 1972).

Readily utilisable nitrogen sources can also repress antibiotic production. For example, the biosynthesis of cephalosporin in *S. clavuligerus* has been shown to be depressed by ammonium (Bascaran *et al.*, 1989a; 1989b). The production of chloramphenicol by *S. venezuelae* has been studied with regard to the influence of nutrient conditions prevailing at the onset of secondary metabolism. Under conditions of nitrogen limitation, the activity of ESM was shown to depend on the concentration of glucose (Bhatnager *et al.*, 1988). Conversely, the extent to which glucose suppressed the production of cephamycin depended on the residual nitrogen source.

All of the above examples are consistent with the proposal that integrated sensory systems monitor the level of carbon, nitrogen and phosphate in the growth medium and may have a cumulative effect on the decision to produce antibiotics and undergo morphological differentiation. Although the molecular mechanism(s) which mediate the carbon and nitrogen regulation of secondary metabolism have not been described, it is presumed that the binding of ancillary proteins to DNA around the sites of transcription initiation will be important.

### 1.8 Role of "Signal" Molecules in Controlling Development

As described in the previous section, the principal factor controlling the onset of antibiotic biosynthesis is probably the deficiency of one or more nutrients required for growth. Recently, the possible role of an "endogenous" signal molecule, guanosine 3'-

diphosphate 5'-diphosphate (ppGpp), in mediating the switch to antibiotic production has been investigated. In *E. coli*, ppGpp mediates the stringent response, brought about by amino acid starvation (Cashel, 1969; 1975). The depletion of amino acids in the growth medium leads to a rise in the intracellular concentration of ppGpp. In turn, this results in a rapid decrease in the rates of accumulation of tRNA and rRNA, due to changes in the transcription rate and promoter specificity of RNA polymerase. A similar type of response has also been demonstrated in streptomycetes (Ochi, 1986). Relaxed mutants (unable to elicit the stringent response) of various *Streptomyces* spp., for example *S. coelicolor* and *S. griseus*, are also impaired in antibiotic production and have reduced levels of sporulation (Ochi, 1990a and 1990b). Recently, studies on *S. coelicolor* A3(2) have shown that ppGpp could not be detected in the early and mid exponential growth phases, but increased levels were observed at the end of exponential growth (Strauch *et al.*, 1990). The *actIII* gene, which is involved in actinorhodin production, was transcribed immediately after the culture entered stationary phase. After nutrient shiftdown, a switch to secondary metabolism was demonstrated by the appearance of the *actIII* transcript one hour later. It was proposed that ppGpp, along with other factors, plays a major role in determining the onset of secondary metabolism (Strauch *et al.*, 1990). It would be interesting to establish how the expression of *bldA* responds to changes in the guanosine nucleotide pools, since in contrast to the general decrease in the levels of other tRNA transcripts, an increase in the abundance of *bldA* transcripts would be expected prior to secondary metabolism (Chater, 1989). In *E. coli*, there is some evidence for the positive regulation of some genes during the stringent response (Cashel and Rudd, 1987).

Another interesting example of a "signal" molecule which controls secondary metabolism and morphological differentiation in some *Streptomyces* spp. is A-factor (2-(6'-methylheptanoyl)-3R-hydroxymethyl-4-butanolide). In *S. griseus*, A-factor induces sporulation, streptomycin biosynthesis and resistance, even at extremely low concentrations ( $10^{-9}$  M) (Hara and Beppu, 1982a and 1982b). This factor may have a key role in synchronising antibiotic production and resistance by preventing

hyphal filaments, whose growth is not yet limited by nutrient depletion, being killed by streptomycin produced from other regions of the colony which have reached the end of vegetative growth. A protein has been isolated that binds A-factor in *S. griseus* (Mikaye *et al.*, 1989). Mutants of *S. griseus* that are deficient in A-factor-binding protein do not require A-factor for streptomycin biosynthesis or sporulation. Indeed, the mutants undergo morphological and physiological development at an earlier stage of growth than the wild type. It was proposed that A-factor-binding protein is a repressor-type regulator which binds to regulatory regions within some or all of the genes for streptomycin biosynthesis, blocking transcription. A-factor is assumed to interact with the repressor protein, rendering it unable to bind DNA and thus, inducing transcription (Miyake *et al.*, 1990).

A-factor analogues, with a butanolide ring as the nucleus of their structure, have also been reported, for example I-factor controlling anthracycline biosynthesis in *S. viridochromogenes* (Gräfe *et al.*, 1982), a factor controlling cytodifferentiation and anthracycline production in *S. bikiniensis* and *S. cyaneofuscatus* (Gräfe *et al.*, 1983) and "inducing factors" controlling virginiamycin production in *S. virginiae* (Nihara *et al.*, 1988). Recently, a protein has been detected which binds the "inducing factor" of *S. virginiae* (cited in Miyake *et al.*, 1990).

A-factor production was shown to be distributed widely among the streptomycetes (Hara and Beppu, 1982). The role of A-factor is not obvious for most of these species, other than the streptomycin-producing *S. griseus* and *S. bikiniensis*. A-factor-deficient mutants of *S. coelicolor* A3(2) and *S. lividans* showed no detectable phenotype (Horinouchi *et al.*, 1986 and 1987). Interestingly, no binding protein specific for A-factor could be detected in either of these species (Miyake *et al.*, 1990). It is unlikely, therefore, that A-factor has a role in controlling the development of these *Streptomyces* spp.

### 1.9 RNA Polymerase Heterogeneity in *Streptomyces*

The controlled expression of alternative sigma factors ( $\sigma$ ), which direct the transcription of different sets of genes, is presumed to have

an integral role in coordinating both the morphological and physiological development of *Streptomyces*. Alternative sigma factors and thus, alternative RNA polymerase holoenzymes ( $E\sigma$ ) have a key role in the control of sporulation in *B. subtilis* (Losick and Pero, 1981; Pero *et al.*, 1982), T4 phage development in *E. coli* (Kassavetis and Geiduschek, 1984), nitrogen utilisation in enteric bacteria (Hirshman *et al.*, 1985; Hunt and Magasanik, 1985), the production of heat shock proteins in *E. coli* (Grossman *et al.*, 1984; Grossman *et al.*, 1986) and the expression of flagellar and chemotaxis genes in *E. coli*, *Salmonella typhimurium* and *B. subtilis* (Helmann and Chamberlain, 1987; Chamberlain, pers. comm.). Recently, the development of spores from aerial hyphae in *S. coelicolor* has been shown to be under the control of an alternative sigma factor,  $\sigma^{whiG}$  (Chater *et al.*, 1989). Interestingly the amino acid sequence of  $\sigma^{whiG}$  closely resembles  $\sigma^{28}$  of *B. subtilis*, the sigma factor involved in directing expression of the chemotaxis and motility regulons. This suggested to Chater and colleagues that  $E\sigma^{whiG}$  may make contact with promoters whose -35 and -10 regions conform to the consensus sequence derived from promoters recognised by  $E\sigma^{28}$  from *B. subtilis*. In support of this hypothesis, the cloning of a DNA fragment containing a  $\sigma^{28}$ -dependent promoter from *B. subtilis*, on a high-copy-number plasmid, caused reduced sporulation in *S. coelicolor* (Chater *et al.*, 1989).

RNA polymerase heterogeneity was first demonstrated in *S. coelicolor* using the *veg* and *ctc* promoters from *B. subtilis* as substrates in *in vitro* transcription "run-off" experiments (Westpheling *et al.*, 1985). The *veg* promoter belongs to the major eubacterial class, which is recognised by  $E\sigma^{43}$  in *B. subtilis*, while the *ctc* promoter is transcribed by a minor form of RNA polymerase holoenzyme,  $E\sigma^{37}$ . Fractionation of crude RNA polymerase preparations from *S. coelicolor* permitted the isolation of two distinct forms of RNA polymerase,  $E\sigma^{35}$  and  $E\sigma^{49}$ , that recognised *veg* and *ctc* respectively. The *ctc*-transcribing holoenzyme, but not the *veg* form, was also shown to transcribe the *endoH* gene (encoding a secreted endonuclease) from *S. plicatus*. The *endoH* promoter resembles the consensus for *B. subtilis* promoters recognised by  $E\sigma^{37}$ .

The *dagA* gene (encoding an extracellular agarase gene) is transcribed by four promoters, whose sequences are markedly heterologous (Buttner *et al.*, 1987). Using the *dagA* gene as a template in *in vitro* transcription assays, three different transcribing activities were resolved. Each RNA polymerase holoenzyme could transcribe only one of the four promoters (Buttner *et al.*, 1988). The  $E\sigma^{49}$ , which was identified previously and recognises *ctc*, transcribes from the *dagAp3* promoter, whereas a novel species,  $E\sigma^{28}$ , recognises the *dagAp2* promoter. Circumstantial evidence suggested that the third holoenzyme, which recognises the *dagAp4* promoter, is  $E\sigma^{35}$  (*veg*-transcribing holoenzyme) which was identified previously. Consistent with this proposal, the promoter sequence of *dagAp4* resembles closely the consensus for the major class of eubacterial promoters. The *dagAp1* promoter could not be transcribed by any of the purified holoenzymes and so, the authors suggested that another form of RNA polymerase must recognise this promoter (Buttner *et al.*, 1988). It is also possible that ancillary proteins are required to activate transcription from *dagAp1*. The labile nature of the *dagAp1*-transcribing activity during purification is consistent with this proposal. The galactose operon of *S. coelicolor* is transcribed from two promoters (Westpheling and Brawner, 1989). Transcription from *galp1* is repressed by growth on glucose and is induced to high levels by growth on galactose. Constitutive transcription from *galp2* occurs at low levels. The *galp1* and *galp2* promoters are thought to be transcribed by  $E\sigma^{28}$  and  $E\sigma^{35}$  respectively.

Using an oligonucleotide probe, designed from a sequence of 10 amino acids that are conserved completely between  $\sigma^{43}$  of *B. subtilis* and  $\sigma^{70}$  of *E. coli* (the "housekeeping" sigma factors encoded by *rpoD*), four genes from *S. coelicolor* have been cloned and sequenced (Tanaka *et al.*, 1988). Each gene encodes a predicted product with high sequence similarity in the regions conserved between  $\sigma^{70}$  of *E. coli* and  $\sigma^{43}$  of *B. subtilis* (*rpoD* boxes). These genes have been designated *hrdA*, *hrdB*, *hrdC* and *hrdD* (homologous to *rpoD*). Independently, the *rpoD* gene of *Myxococcus xanthus* was used also to clone *hrdB*, *hrdC* and *hrdD* (Buttner *et al.*, 1990). The degree of similarity between the predicted gene

products of the *hrd* genes and those of the *rpoD* genes of *E. coli* and *B. subtilis*, far exceeds that observed between different  $\sigma$  factors for minor classes of promoters (Tanaka *et al.*, 1988).

Gene disruption experiments have shown that *hrdA*, *hrdC* and *hrdD* mutants are viable and are unaffected apparently in differentiation, gross morphology and antibiotic production (Buttner *et al.*, 1990; Takahashi *et al.*, cited in Buttner *et al.*, 1990). The inability to isolate *hrdB* mutants using an integration system based on the actinophage  $\phi$ C31, suggested that the  $\sigma$  factor encoded by this gene may be essential for the transcription of the general "housekeeping" genes. Furthermore, it was also noted that the putative  $\sigma$  factors encoded by *hrdA*, *hrdC* and *hrdD* appeared to be unable to complement the transcribing activity encoded by *hrdB* (Buttner *et al.*, 1990).

It has been proposed that closely-related families of  $\sigma$  factors may recognise distinct subsets of promoters with only minor differences in the sequence or spacing of their "-35" and "-10" boxes (Westpheling and Brawner, 1989). Although RNA polymerase heterogeneity has been demonstrated in *Streptomyces*, disappointingly little is known about the discrete gene sets which are recognised by the alternative sigma factors.

### 1.10 Polyketide biosynthesis

The diverse range of secondary metabolites produced by the *Streptomyces* can be classified on the basis of their biosynthetic origins (Turner and Aldridge, 1983). The following products of primary metabolism were identified as being important precursors for secondary metabolites; (i) small carboxylic acids, (ii) amino acids and intermediates of the shikimate pathway, (iii) nucleosides and (iv) intermediates of the tricarboxylic acid cycle.

The tetracyclines, along with the anthracyclines, macrolides, polyethers, ansamycins, polyenes and isochromanquinones are members of the polyketide class of antibiotics, which are synthesized from small carboxylic acids, in a manner analogous to fatty acid biosynthesis. The

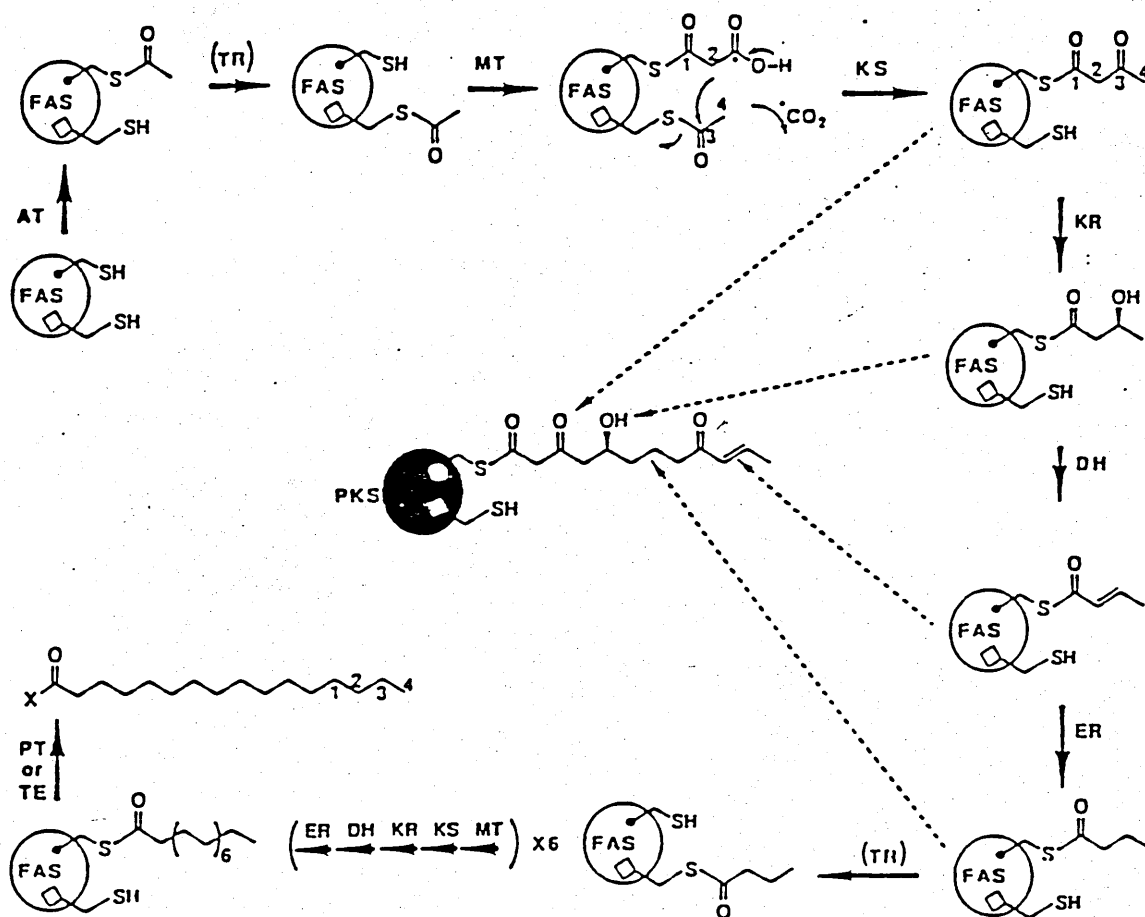


Figure 1.6. Schematic representation of the biosynthesis of fatty acids and polyketides. The circle labelled FAS or PKS represent the fatty acid synthases or polyketide synthases. These complexes carry two thiol groups, one on the  $\beta$ -ketoacyl synthase ( $\square$ ) and the other on the acyl carrier protein (O). Also labelled are the different functions required for each step; acetyl transferase (AT), acyl transfer reaction (TR, which is not unambiguously assigned to a specific component of the complex), malonyl transferase (MT),  $\beta$ -ketoacyl synthase (KS), ketoreductase (KR) dehydratase (DH), enoyl reductase (ER), palmityl transferase (PT) or thioesterase (TE) involved in the release of palmityl CoA ( $X=\text{CoA}$ ) or free palmitic acid ( $X=\text{OH}$ ) respectively. From Hopwood and Sherman, 1990.

similarity of these reactions was demonstrated convincingly by the inhibition of both these pathways using the antibiotic cerulenin (Arison and Omura, 1974). Recently, DNA regions known to be involved in the earliest stages in polyketide biosynthesis have been sequenced. The predicted protein products of these sequences have compelling homology to the  $\beta$ -ketoacyl synthase,  $\beta$ -ketoacyl reductase and acyl carrier protein (ACP) of fatty acid biosynthesis (for review, see Hopwood and Sherman, 1990).

The synthesis of the fatty acid palmitate will be considered (Figure 1.6) to provide a basis for the discussion of polyketide assembly. Palmitic acid is an important precursor for the synthesis of phosphoglycerides, which are the polar lipids found almost exclusively in biological membranes. This fatty acid is a fully-saturated 16 carbon polyketide which is synthesized from one acetyl-CoA unit and seven malonyl-CoA units, by a complex with eight enzymatic activities, called the fatty acid synthase (FAS). The malonyl-CoA extender unit is formed from acetyl-CoA and CO<sub>2</sub> by the action of acetyl-CoA carboxylase. The starter acetyl group is transferred, by the action of an acetyl transferase, to the ACP which in turn transfers the acetyl group to the active site of the  $\beta$ -ketoacyl synthase. Condensation then ensues between the acetyl starter unit and a malonyl extender unit, which is attached to the ACP vacated by the acetyl group, resulting in the loss of CO<sub>2</sub> (the same CO<sub>2</sub> molecule used to produce malonyl-CoA from acetyl-CoA). The ACP transfers the growing polyketide chain, in strict sequence, to the active sites of ketoreductase, dehydratase and enoyl reductase where the keto group from the malonyl CoA extender unit is converted first to a hydroxyl group, then to a enoyl group and finally, to an alkyl group. After this sequence of events the growing polyketide chain is transferred back to the active site of  $\beta$ -ketoacyl synthase. Only now can the second malonyl-CoA extender unit attach to the ACP and participate in a further round of condensation-reduction-dehydration-reduction. The remaining malonyl-CoA units are added in a reiterant cycle of reactions to form palmityl-ACP. A palmityl transferase then releases the palmitic acid from the FAS complex. Therefore, a typical FAS is a multivalent system involving eight functional units;



acetyl, malonyl and palmityl transferases, ACP, ketoacyl synthase, keto reductase, dehydratase and enoyl reductase. The fatty acid synthase that catalyses these steps varies in different life forms. In most bacteria and plant plastids, the multienzyme complex is formed from eight separate polypeptides (Type II FAS). In contrast, the same eight functions are carried out by domains on a single, multifunctional polypeptide (Type I FAS) in invertebrates. The yeast, *Saccharomyces cerevisiae*, represents an intermediate case in which the FAS consists of two polypeptides, carrying three and five catalytic activities respectively.

Despite the obvious similarity in the chemistry of the biosynthesis of polyketide backbones and fatty acids, PKSs appear to be able to introduce complexity at five different levels. (i) The cycle of reduction-dehydration-reduction that follows the addition of two carbons in fatty acid biosynthesis is omitted or curtailed at some or all of the points in polyketide assembly. Thus, the PKS has to choose between four possibilities at each step in assembly and depending on the decision, leaves alkyl, enoyl, hydroxyl or keto groups at different carbon atoms (the occurrence of keto groups on the carbon backbone gives rise to the name "polyketide"). As in fatty acid biosynthesis, the modification of the keto groups probably takes place immediately after condensation, rather than after the complete backbone has been formed. This has been suggested by the incorporation of intermediates, predicted from a FAS-like progressive pathway, into the small polyketide aspyrolle in *Aspergillus*. Putative intermediates that contained functional groups, which could not be predicted from a FAS-type progression, were not incorporated (Staunton, pers. comm.). Furthermore, intermediates in the assembly of the polyketide backbone of the mycinamicins and tylosin have been isolated, which show the pattern of reduction, dehydration, reduction and stereochemistry found in the final product (Kinoshita *et al.*, 1988 and Huber *et al.*, 1990, cited in Hopwood and Sherman, 1990). (ii) Although polyketide chain initiation is primed commonly with acetyl-CoA, this is not an absolute requirement. For example, *Saccharopolyspora erythraea* uses propionate as a starter unit in the biosynthesis of erythromycin. In contrast to FAS, many PKS make a

choice of extender unit from a range of possibilities including acetate, propionate, butyrate and occasionally more complex alternatives. (iii) As a consequence of the condensation-reduction-dehydration-reduction cycle being broken during the assembly of many polyketides, the carbon chains often contain chiral centres, unlike the products of FAS which are achiral. (iv) The biosynthesis of many polyketides including oxytetracyclines, actinorhodin, granaticin, and tetracenomycin, involves cyclization or aromatisation of the polyketide backbone. The protein products of *act* ORF4 and *gra* ORF 4, are deduced to contain a  $\beta$ -ketoacyl cyclase activity as they can complement *actVII* mutants which accumulate an incorrectly cyclised shunt product, mutactin (Zhang *et al.*, 1990). Due to significant homology to the deduced products of *act* and *gra* ORF4, the predicted products of *whiE<sub>A</sub>*<sup>ORF VI</sup> and *tcm* ORF4 are also thought to have cyclase activity (Hopwood and Sherman, 1990). The later stages of monensin biosynthesis involve epoxidases which generate an intermediate that undergoes a cascade of ring closures, to form the heterocyclic rings seen in this antibiotic (Cane *et al.*, 1983). The three oxygens at positions 7, 8 and 9 (Figure 1.7) of the cyclic ethers, are derived from molecular oxygen by epoxidation of C=C double bonds. The other oxygen atoms, including the two in the remaining cyclic ethers are derived from their respective propionate or acetate extender units. (v) The polyketide structure is modified usually by the addition of side groups to the polyketide nucleus. Unusual sugar motifs are attached often to macrolide antibiotics. For example, the deoxysugars mycarose and mycinose and the deoxy-amino sugar mycaminose, are added to the tylactone ring during the later stages of tylosin biosynthesis (Baltz *et al.*, 1981). The biosynthesis of oxytetracycline (Section 1.14) provides an example of a polyketide nucleus which is modified by the addition of numerous side chains. Several antibiotics have bilateral symmetry and are thought to be produced by dimerisation of two identical polyketide nuclei. Examples include, actinorhodin, produced by *S. coelicolor* (Section 1.11) and chlorothricin, produced by *S. antibioticus* (Lee *et al.*, 1986).

The biosynthesis of the polyether monensin A by *S. cinnamonensis* provides an excellent example of the complex "programming" required to

Carboxylic Acids; PPAPBAPAAPPP  
Group Substitutions; ZWXWXWZYYY

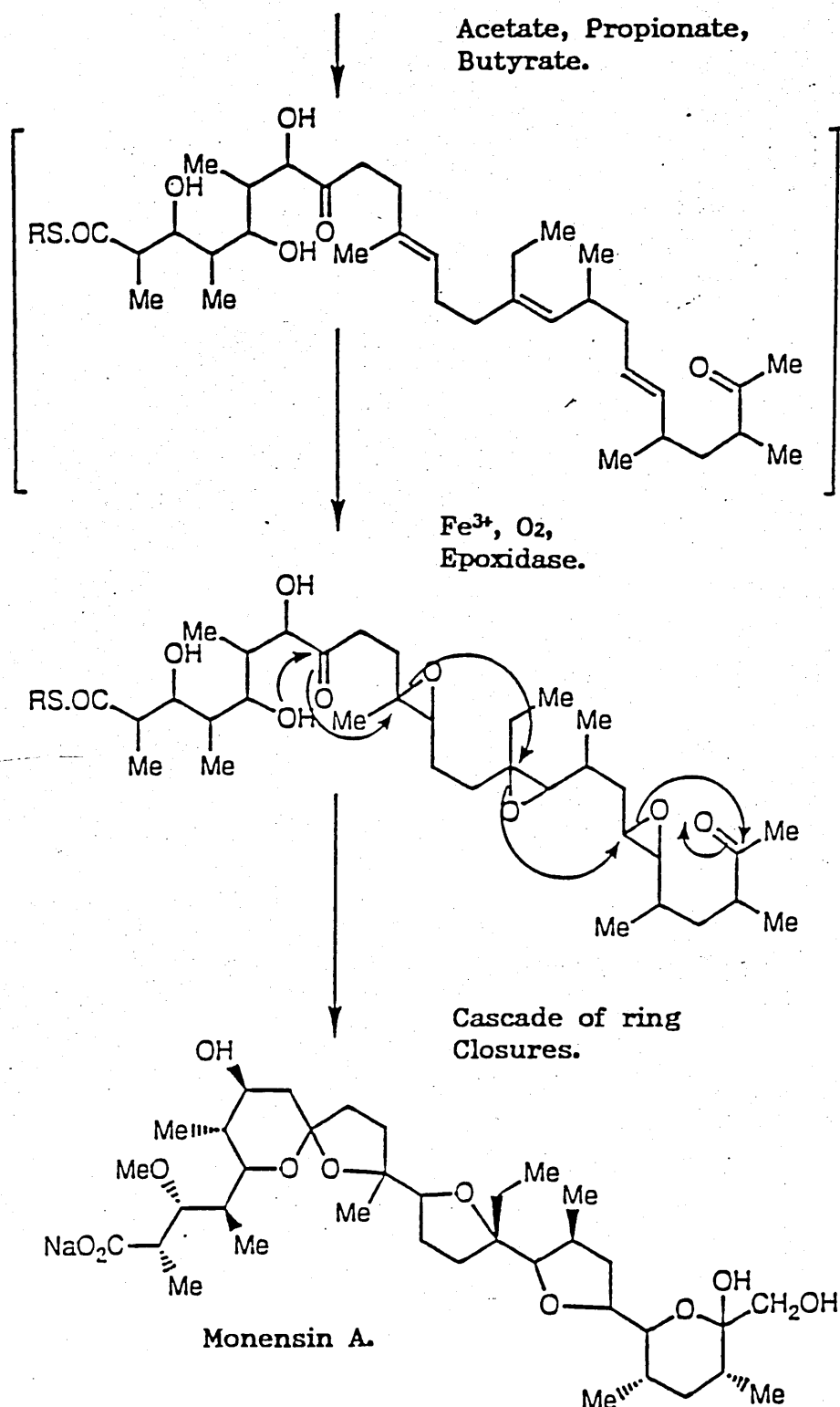


Figure 1.7. The proposed biosynthetic route to monensin A. The hypothetical polyketide skeleton is derived by the condensation of five acetyl (A), seven propionyl (P) and one butyryl (B) residues in the order shown. Also indicated are the polyketide group substitutions; alkyl (W), enoyl (X), hydroxyl (Y) and keto (Z) functionalities. Modified from Cane *et al.*, 1983.

synthesize the polyketide backbone of some antibiotics. During polyketide assembly, the monensin PKS has to make 37 decisions: choices of 13 building units in a specific order from acetate, propionate and butyrate; a choice of four types of chemistry after each of the 12 condensations, to give either keto [Z], hydroxyl [Y], enoyl [X] or alkyl [W] and correct choice of stereochemistry at 12 points in the assembly of the carbon skeleton. The monensin PKS produces only one main polyketide product from the vast number of theoretical possibilities (Figure 1.7).

### 1.11 Actinorhodin Biosynthesis

The amenability of *S. coelicolor* to both *in vivo* and *in vitro* genetic manipulations and the relative ease with which actinorhodin producing colonies can be detected (on alkali medium they are deep blue) have contributed to the biosynthesis of actinorhodin being a paradigm for the chain assembly of other polyketides, including oxytetracycline. The primary carbon skeleton of actinorhodin is derived from 16 acetate units. A series of mutants blocked in actinorhodin biosynthesis (*act* mutants) was isolated and classified into seven phenotypic classes (Rudd and Hopwood, 1979). The *act* loci were shown to be linked on a short segment of the *S. coelicolor* chromosome, by analysing the frequencies of recombinants after conjugal matings between *act* mutants and wild-type cultures. Pairwise cosynthesis between *act* mutants allowed the mutant classes to be arranged in a linear sequence of biosynthetic blocks; I, III, VII, IV, VI and V (Rudd and Hopwood, 1979). The *actII* mutants did not produce actinorhodin in cosynthesis tests with any of the other mutants and were presumed to be deficient in a positive-regulatory factor. The *actI* and *actIII* mutants failed to secrete any biosynthetic intermediate which could be converted to actinorhodin by any of the other mutant classes, but were able to convert intermediates secreted by all other mutant classes to actinorhodin. This cosynthesis data suggested that *actI/III* mutants were defective for enzymes involved in the earliest steps of biosynthesis. The *actIII* mutants accumulated a red pigment, whereas *actI* mutants did not and were deduced, therefore, to be blocked at a later stage in the construction of the carbon skeleton (Rudd and Hopwood, 1979).

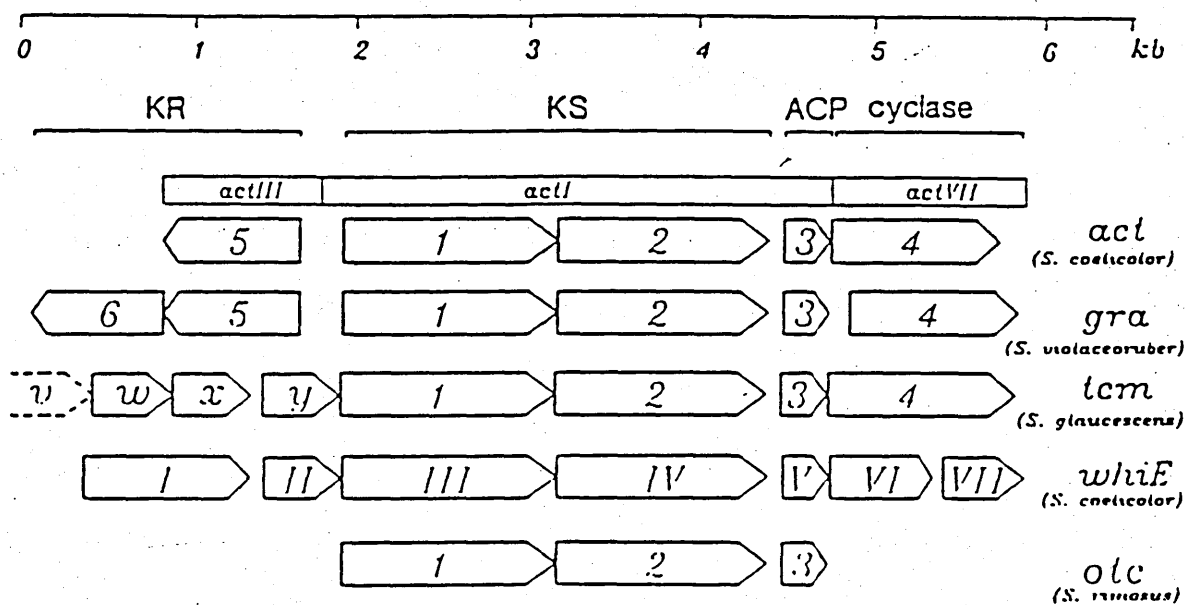


Figure 1.8. Organisation of ORFs in the PKS gene cluster for actinorhodin in *S. coelicolor*, granaticin in *S. violaceoruber*, tetracenomycin in *S. glaucescens*, the putative spore pigment in *S. coelicolor* and oxytetracycline in *S. rimosus*. From Hopwood and Sherman, 1990.

It was suggested that genes or gene clusters responsible for the synthesis of polyketide chains, might have a common evolutionary origin (Malpartida and Hopwood, 1984). Using probes derived from segments of DNA that complemented *actI* and *actIII* mutants respectively, Southern analysis identified homologous fragments in genomic-DNA digests of several other polyketide-producing streptomycetes (Malpartida *et al.*, 1987). Fragments cross-hybridizing with the *actI* probe, were identified in the oxytetracycline producer *S. rimosus*, the tetracenomycin producer *S. glaucescens* and the granaticin producer *S. violaceoruber*. Analysis of the predicted amino acid sequence of the *actIII* gene product revealed that it encodes a reductase (Hallam *et al.*, 1988). No fragment hybridized with the *actIII* probe in genomic DNA digests of *S. glaucescens*. However, this was not unexpected as the biosynthesis of tetracenomycin G does not require the corresponding reduction.

### 1.12 Organisation of Polyketide Synthase Genes

The *actI* and *actIII* regions have been sequenced (Fernandez and Malpartida, cited in Hopwood and Sherman, 1990) along with the *actVII* region, which is colinear with the *actI* region and is thought to have a role in the cyclization of the polyketide chain (Zang *et al.*, 1990). A number of ORFs have been identified, which encode protein products with convincing homology to enzymes involved in fatty acid synthesis. These ORFs have been designated the actinorhodin polyketide synthase (*act* PKS) genes. Sequence data has also been obtained for the PKS genes of granaticin (Sherman *et al.*, 1989), oxytetracycline (Sherman, pers. comm.) and tetracenomycin (Bibb *et al.*, 1989), which cross-hybridized with the *act* probes described above. The sequencing of DNA, which complements mutants of *S. coelicolor* that are deficient in spore pigmentation (*whiE*), has also revealed a set of ORFs that clearly correspond to those of known PKS genes (Davis and Chater, 1990).

The organisation of the *act*, *gra*, *otc*, *tcm* and *whiE* PKS genes has been compared in a recent review (Hopwood and Sherman, 1990). Each PKS gene cluster contains three characteristic ORFs, designated ORF 1-2-3 (except for the *whiE* ORFs which are called III-IV-V) (Figure 1.8).

The predicted protein sequences of ORF 1-2 show convincing homology to the fatty acid  $\beta$ -ketoacyl synthase of *E. coli*, encoded by *fabB*. The predicted protein sequence of ORF 3 in each cluster resembles strongly acyl carrier proteins (ACPs) from FAS in a variety of life forms, including bacteria, plants and animals. Immediately downstream of ORF 1-2-3 there is a fourth ORF that shows relatedness between different PKS clusters.

In contrast with the predicted products of ORF 1-2-3, the presumed products of the fourth ORFs only resemble each other in a specific domain. The *act* and *gra* ORF4 sequences show homology throughout their entire length. These particular ORFs are both thought to encode a bifunctional cyclase/dehydratase, as they complement *actVII* mutants, which accumulate mutactin and are thought to be defective in cyclization and an accompanying dehydration. The predicted product of *whiE* ORF VI, resembles only the amino-terminal half of the *act* and *gra* ORF4 proteins. It was deduced, therefore, that the *whiE* ORF VI may encode a monofunctional polyketide cyclase (Hopwood and Sherman, 1990). The *tcmIa* ORF4 protein was thought originally to be a bifunctional cyclase/O-methyltransferase. However, recent evidence (described in Chapter 4), suggests that there may be an error in the sequence and that the *tcm* PKS, similar to *whiE*, may actually contain a monofunctional cyclase (Hutchinson, pers. comm.). The DNA sequence of the ORF4 region of *S. rimosus* has not been completed (Sherman, pers. comm.). However, a region of DNA immediately downstream of *otc* ORF1-2-3 does hybridize with *actVII* DNA (Butler *et al.*, 1990).

The *actIII* gene (*act* ORF5), which is encoded on the opposite strand to ORF 1-2-3, is predicted to encode a ketoreductase (Hallam *et al.*, 1988). Recently, the introduction of a plasmid carrying only the intact *actIII* (*act* ORF5) gene into the 2-hydroxyalkavinone producer *Streptomyces galilaeus* ATCC 31671 was shown to result in the exclusive formation of the hybrid antibiotic alkavinone (Bartel *et al.*, 1990). The chemical analysis of alkavinone and other hybrid polyketides produced by *S. galilaeus*, when transformed with DNA from *S. coelicolor* containing different combinations of the *actI*, *III*, *IV* and *VII* loci, suggested that

the function of the *act* ORF5 product was to reduce the keto group at C-9 from the carboxy terminus to a hydroxyl group (Bartel *et al*, 1990). This proposal is consistent with the finding that the predicted sequence of the *act* ORF5 product has significant homology to other reductases (Hallam *et al*, 1988). The predicted amino acid sequence of the *gra* ORF5 and ORF6 proteins show high similarity to the predicted *actIII* gene product and possibly encodes a heterodimeric ketoreductase (Hopwood and Sherman, 1990). Sequencing of a fragment within the *otc* cluster, which cross-hybridized with *actIII* DNA (Butler *et al*, 1990), has revealed a potential protein coding region that is predicted to encode a product with high amino acid sequence identity to the deduced *actIII* product (Thamchaipenet, pers. comm.). The regions in the *otc* cluster, which hybridized to *actI* and *actIII* DNA, are separated by approximately 8 kb (Butler *et al*, 1990) indicating that the genetic organisation of the *otc* PKS is topologically different to the other PKS gene clusters so far analysed.

### 1.13 The Module Hypothesis

The biosynthesis of the macrolide antibiotic erythromycin A (erA) by *Saccharopolyspora erythrae*, is a complex pathway that includes; (i) the formation of a 14-membered macrolactone ring from the condensation of propionyl-CoA and 6 methylmalonyl-CoA residues, (ii) the synthesis and attachment of sugar residues to form erythromycin D (erD) and (iii) the hydroxylation and O-methylation of erD to form erA. Genes involved in the synthesis of the macrolactone ring are designated *eryA* and are thought to be closely linked in a 30 kb region of DNA. Cross-hybridization between different fragments within the 30kb segment and limited sequencing suggested that the *eryA* region was organised into six repeated modules of approximately 5 kb, each containing an acetyl transferase, ACP,  $\beta$ -ketosynthase and  $\beta$ -ketoreductase (Katz *et al*, 1990). As the assembly of erythromycin requires six condensations, it was proposed that each module contains the enzymatic activities required for the condensation and processing of only one of the six methylmalonyl-CoA extender units (Katz *et al*, 1990).

Recently, a 10kb segment of DNA from the *eryA* region, which was



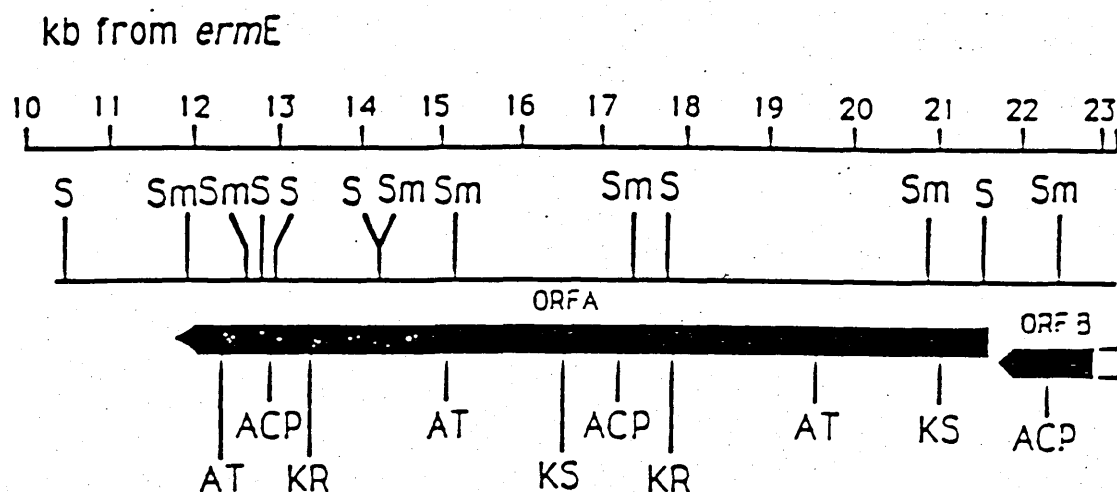


Figure 1.9. Potential catalytic sites in the deduced protein of ORFA in the *eryA* region of the *Saccharopolyspora erythraea*: ACP, acyl carrier protein; KS,  $\beta$ -ketoacyl-ACP synthase; AT, acyltransferase or thioesterase and KR,  $\beta$ -ketoacyl-ACP reductase. Abbreviations: S, *SacI*; Sm, *SmaI*. Modified from Cortes *et al.* (1990).

cloned by chromosome walking from the *ermE* resistance determinant, has been sequenced completely (Cortes *et al.*, 1990). Analysis of the sequence revealed a PPCR extending over 9.5kb. The deduced protein product contains nine separate portions, whose amino acid sequences are very similar to the amino acid sequences for the catalytic domains of known fatty acid and polyketide synthases (Figure 1.9). The N-terminal and C-terminal halves of the deduced protein each contain a set of four predicted activities in the same order, except for the presence of an additional acyltransferase (or thioesterase) domain at the C-terminus. It would appear that a single, multifunctional peptide, which resembles type II FAS in invertebrates, might catalyse more than two of the six possible condensation cycles. In contrast, relatively simple polyketide structures such as actinorhodin, granaticin, tetracenomycin and oxytetracycline would appear to be synthesised by multienzyme complexes, which contain single polypeptides for each catalytic activity (Hopwood and Sherman, 1990).

It was suggested that the large number of genes involved in the assembly of the macrolactone ring may reflect the complexity of the structure (Katz *et al.*, 1990). It is conceivable that other macrolide antibiotics, including tylosin, avermectin and perhaps polyethers such as monensin will also require a number of PKS modules for assembly of the carbon chain. The sequencing of many of these clusters is currently underway and should prove enlightening in the near future.

#### 1.14 Biosynthesis of Oxytetracycline

Oxytetracycline (5-hydroxytetracycline; OTC) is one of the commercially-important broad-spectrum, tetracycline series of polyketide antibiotics produced in large-scale fermentations by *S. rimosus*. Within the tetracycline group, deoxycycline (6-deoxy-hydroxytetracycline), which is made by a chemical modification of OTC and chlorotetracycline (N-demethyl-7-chlorotetracycline; CTC), produced by *S. aureofaciens*, also have important pharmaceutical properties. A large number of tetracyclines have been reported, some of which are made by microorganisms other than streptomycetes, for example, *Nocardia* (Sinclair *et al.*, 1962).

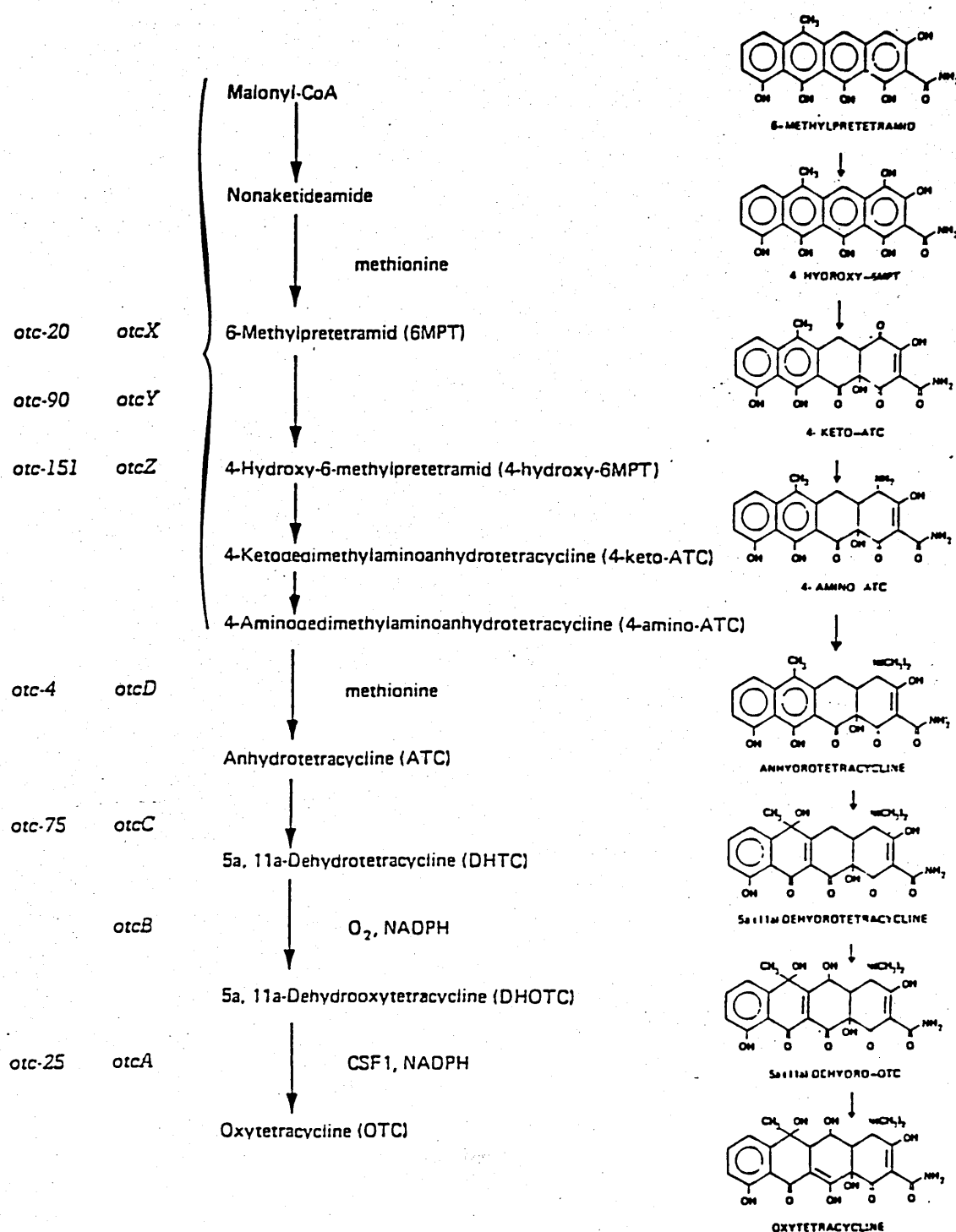


Figure 1.10. Biosynthesis of oxytetracycline. Indicated are the proposed steps at which various OTC-negative mutants are blocked. From Rhodes *et al.*, 1981.

The plausible biosynthetic pathway for the tetracyclines was elucidated at the Lederle Laboratories (McCormick, 1966; McCormick and Jensen, 1969) by the biochemical analysis of intermediates produced by mutants blocked in production. Confirmation of the polyketide nature of the carbon skeleton was obtained by NMR spectroscopy of OTC, labelled with  $^{13}\text{C}$ -labelled acetate and malonate (Thomas and Williams, 1983).

Chemical considerations predict at least nine stages leading to oxytetracycline (Figure 1.10) (Rhodes *et al.*, 1981); (i) the sequential condensation of eight malonyl-CoA extender units onto a malonamyl-CoA primer unit to produce a nonaketide and then cyclization to form the tetracycline nucleus, (ii) C-6 methylation to form the first isolatable intermediate 6-methylpretetramid (6MPT), (iii) C-4 hydroxylation to form 4-hydroxy-6MPT, (iv) C-6 oxidation to form 4-keto-anhydrotetracycline (4-amino-ATC), (v) C-4 transamination to form 4-keto-ATC, (vi) double N-methylation to form ATC, (vii) C-6 hydroxylation to form dehydrotetracycline (DHTC), (viii) C-5 hydroxylation to form dehydrooxytetracycline (DHOTC) and finally (ix) CSF1-dependent reduction to form OTC.

Genetic analysis of mutants blocked in OTC production indicated that the oxytetracycline (*otc*) genes were located on the chromosome of *S. rimosus* (Pigac and Alacevic, 1979). The first integrated study of the genetics and biochemistry of OTC production was undertaken at the Pfizer Laboratories (Rhodes *et al.*, 1981). A series of mutants blocked in OTC production was isolated and classified primarily on the basis of cosynthesis tests in liquid culture, into nine groups. Single representatives from each group (*otc*-4, -19, -20, -25, -56, -75, -90, -151, and -155) were selected for further analysis. Three of these mutants, *otc*-19, -56 and -155, were able to produce OTC when supplied with culture filtrate of the Coenzyme F<sub>420</sub>-producing strain *S. aureofaciens* ATCC 13190. This suggested that these *otc* mutants were blocked in the synthesis of this cosynthetic factor, which is essential for the final enzymatic step in OTC biosynthesis. The remaining six classes of *otc* mutants were presumed to be blocked at six different steps in the biosynthetic sequence from malonyl-CoA to oxytetracycline.

The absolute order in which the *otc-4*, *-20*, *-25*, *-75*, *-90* and *-151* mutants were blocked could not be deduced, as each representative failed to cosynthesize OTC with the other *otc* mutants in agar strip test in any of the possible pairwise combinations (as described by Delic *et al.*, 1969). The feeding of known intermediates of OTC did, to some extent, delineate the steps which could be blocked (Rhodes *et al.*, 1981). All of the *otc* mutants, except *otc-25*, could convert a derivative of DHTC to OTC. It was argued that the *otc-25* mutant could not be blocked in C-5 hydroxylation, as dehydroxytetracycline would probably be converted to tetracycline, which is biologically active: the *otc-25* mutant has an antibiotic-negative phenotype. The *otc-25* mutant was presumed, therefore, to be blocked in the reduction of DHOTC to form OTC. The *otc-4* mutant was able to convert ATC, but not 4-amino-ATC, to OTC and was deduced, therefore, to be blocked in N-methylation of 4-amino-ATC. Similarly, the *otc-75* mutant could not convert ATC to OTC and was thought consequently to be blocked in the C-6 hydroxylation that forms DHTC. Having assigned the *otc-25*, *-4* and *-75* mutants to the steps between 4-amino-ATC, which would produce antibiotic-negative phenotypes if disrupted, it was concluded that *otc-20*, *-90* and *-151* were blocked before the formation of 4-amino-ATC. Mutant in the class represented by *otc-90* were pale and nonpigmented. As many of the intermediates and shunt products of the pathway are highly coloured, it was considered probable that the *otc-90* mutant was blocked in the earliest biosynthetic steps (Hunter, pers. comm.). The mutant classes represented by *otc-4*, *-20*, *-25*, *-75*, *-90*, and *-151* were assigned to putative loci *otcD*, *otcX*, *otcA*, *otcC*, *otcY* and *otcZ* respectively (Rhodes *et al.*, 1981). The conversion of DHTC to DHOTC was attributed to a putative locus called *otcB*.

Genetic mapping, using the representatives from each mutant group, suggested that the *otc* genes were located in two distinct clusters (Rhodes *et al.*, 1981). The first (located at 4 o'clock on the circular map of the *S. rimosus* chromosome), was responsible for the formation of the tetracycline nucleus and its conversion to anhydrotetracycline (ATC), while the second (located at 10 o'clock),

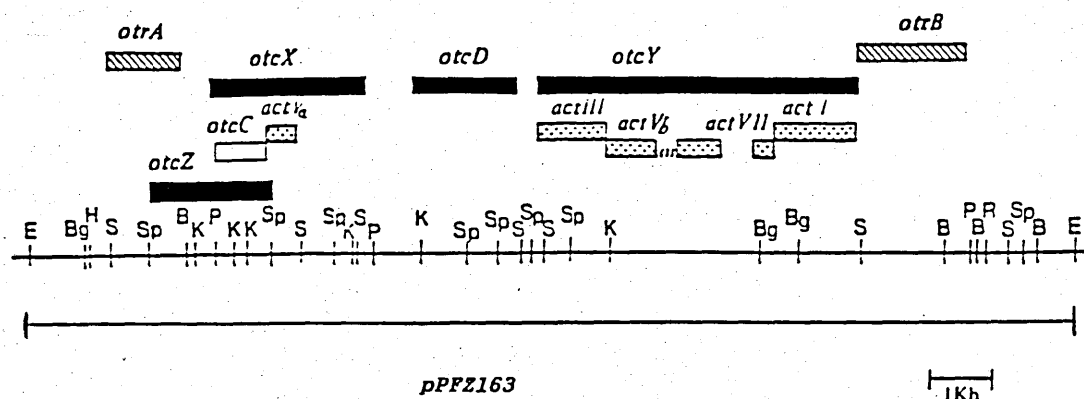


Figure 1.11. Restriction map of the *otc* cluster. Hatched boxes (▨) denote the location of the *otr* resistance genes. Closed boxes (■) show the minimum DNA segments that can complement the OTC-negative classes of mutants (Butler *et al.*, 1989). Stippled boxes (▤) display homology to *act* DNA (Butler *et al.*, 1990). The open box (□) denotes the position of the *otcC* gene located by "reverse genetics" (Binnie *et al.*, 1989). The line (—) delineates the region of DNA in pPFZ163 which confers the ability to produce OTC when introduced in *S. lividans* and *S. albus* (Binnie *et al.*, 1989). Abbreviations: B, *Bam*HI; Bg, *Bgl*II; E, *Eco*RI; H, *Hind*III; K, *Kpn*I; P, *Pst*I; R, *Eco*RV; S, *Sac*I; Sp, *Sph*I. Modified from Butler *et al.*, 1990.

contained two sets of genes whose products either had an enzymatic role in the conversion of ATC to OTC or were involved in the biosynthesis of Coenzyme F<sub>420</sub> (also called CSF<sub>1</sub>).

Initially, the early steps encoded in the 4 o'clock were the focus of attention at the Pfizer Laboratories. Genomic DNA fragments of *S. rimosus*, which conferred resistance to OTC and complemented all of the *otc* mutants (*otc*-4, -20, -90 and -151) blocked in the "early" biosynthetic sequence from malonyl CoA to anhydrotetracycline (ATC), were cloned. The cloned DNA was linked physically within a 35 kb segment of the chromosome. The gene cluster is flanked by two resistance genes, which can independently confer protection against the antibiotic (Figure 1.11) (Butler *et al.*, 1989).

At the Pfizer laboratories, attention turned subsequently to the functions involved in the conversion of ATC to OTC. The product of the *otcC* gene, anhydrotetracycline (ATC) oxygenase, has an easily assayable enzymatic activity. A "reverse genetic" approach was taken to clone the *otcC* gene and linked DNA, that is the 10 o'clock cluster. The ATC oxygenase was purified and the N-terminal sequence determined. Oligonucleotide probes were designed and used to clone a corresponding fragment of chromosomal DNA from *S. rimosus*. The screening of a cosmid library, using a probe derived from the cloned DNA, located the *otcC* gene within the cluster for the early steps in OTC production and not as expected (Rhodes *et al.*, 1981) at a separate locus on the opposite side of the chromosomal map (Binnie *et al.*, 1989). The hybridization results contradicted the original genetic mapping data and it became a formal possibility that all the structural genes for OTC biosynthesis were located in the 4 o'clock cluster. The ability of a plasmid carrying this gene cluster (on a 34 kb *EcoRI* fragment) to confer OTC production when introduced into the heterologous hosts *S. albus* and *S. lividans*, demonstrated convincingly that all the structural genes for OTC were indeed linked on a single segment of the *S. rimosus* chromosome (Binnie *et al.*, 1989).

Recently, it has been shown that the ATC oxygenase also has a requirement for a flavin cosynthetic factor, which is distinct from CSF1 used by the tetracycline dehydrogenases (Vancurova *et al.*, 1988). Although the *otc-75* mutant is blocked clearly in the conversion of ATC to DHTC, it would now appear that the mutation, which was mapped outside the 4 o'clock cluster, disrupts the biosynthesis of a cofactor required for ATC oxygenase activity and not the expression of the *otcC* gene product itself. There is also no convincing mapping data available to support the published location of the *otc-25* mutation (Rhodes *et al.*, 1981) to the 10 o'clock cluster (Binnie *et al.*, 1989). The heterologous production of OTC, however, demonstrates clearly that the *otcA* locus is located within the 34 kb *EcoRI* fragment, flanked by the two resistance genes. Attempts are underway to locate the *otcA* locus by complementation of the *otc-25* mutant (Binnie, pers. comm.).

Using a series of overlapping DNA fragments from the actinorhodin and oxytetracycline clusters, it has been possible to define regions which cross hybridize (Butler *et al.*, 1990). Substantial hybridization signals were observed between the following regions (Figure 1.11). (i) The *actI* and *actIII* DNA probes both hybridized to the DNA segment that complements the *otcY* mutants, which is consistent with the previous proposal (Hunter, pers. comm.) that the *otcY* mutants were blocked in the earliest steps in biosynthesis. (ii) An *otc* DNA fragment adjoining the *actI*-hybridizing region hybridizes to *actVII* DNA. This suggests that there could be a gene encoding a putative cyclase in the *otc* cluster, downstream of the genes for the  $\beta$ -ketoacyl synthase and ACP, as found in the *act*, *gra*, *tcm* and *whiE* PKS clusters (Section 1.12). (iii) Part of the *actVa* region, which has been implicated in hydroxylation (Hopwood *et al.*, 1985; Cole *et al.*, 1987) hybridizes to a segment of the DNA that complements *otcX* mutants. The C-4 hydroxylation of 6-methylpretetramid is one of the possible steps, which could be blocked in the *otcX20* mutant. The cross-hybridization between the *actVa* region and the *otcX* region might indicate a segment of DNA required for hydroxylation. (iv) A region of *actVB* DNA hybridizes to a segment located between the regions that hybridize to *actI* and *actIII*. It has been proposed that mutations in the *actVB* region could be



defective in a hydroxylation step, which is separate from *actVa* hydroxylation (Cole *et al.*, 1987). In the biosynthesis of OTC, the conversion of dehydrotetracycline to dehydroxytetracycline requires a C-5 hydroxylation. Should the region of *actVB*, which hybridizes to the *otc* cluster, contain a structural gene involved in hydroxylation, it is possible that the region of the *otc* cluster that cross-hybridizes with *actVB* DNA could encode the *otcB* gene.

The *otrA* resistance gene product (OtrA) protects the ribosomes from translational arrest by OTC, while the product of the *otrB* resistance gene results in decreased accumulation (increased efflux) of the antibiotic from the mycelium (Ohnuki *et al.*, 1985). The *otrA* gene has been cloned from the Pfizer production strain *S. rimosus* M15883 (Rhodes *et al.*, 1984) and sequenced (Doyle, 1987). The predicted amino acid sequence of OtrA is very similar to the deduced amino acid sequence of elongation factor G (EF-G) from *E. coli* (Zengel *et al.*, 1984). It has been proposed that OtrA may function as an alternative EF-G, which substitutes for the usual EF-G that is presumed to be inhibited by tetracycline (Doyle *et al.*, 1991).

Tandem promoters were proposed for the *otrA* resistance gene (Doyle, 1987). The promoter *otrAp1*, which is nearest the putative translation initiation codon, resembles the consensus for the major class of eubacterial promoters. The sequence of the putative *otrAp2* promoter was not determined as it lay just outside the region which was cloned and sequenced. Transcripts predicted to originate from these tandem promoters were detected in RNA isolated from *S. rimosus* M4018, under conditions which did not support antibiotic production. As no structure resembling a rho-independent terminator was detected in the intergenic region upstream of *otrA*, it was suggested further that *otrA* may be transcribed as part of a polycistronic message along with the biosynthetic gene, *otcZ*, which is located immediately upstream and is transcribed from the same strand as *otrA*, during antibiotic biosynthesis (Doyle, 1987). The presence of tandem promoters suggested that the expression of *otrA* could be modulated during the growth cycle by alternative forms of RNA polymerase.

### 1.15 Scope of Present Work

The work presented in this thesis was concerned primarily with transcriptional analysis at the "*otrA*" end of the oxytetracycline cluster. The specific aims were to identify and characterise all of the promoters pertaining to the expression of the *otrA* resistance gene, to determine the temporal activity of these promoters at different stages during the fermentation of oxytetracycline and to analyse the regions of these promoter for any putative regulatory sequences involved in *otrA* expression.

During the course of this project DNA segments that complemented classes of mutants blocked in OTC production (Rhodes *et al.*, 1981) were also sequenced and analysed.

## **CHAPTER 2**

### **MATERIALS AND METHODS**

## INTRODUCTION

This chapter contains the general procedures used in the experiments which were the basis of this thesis. The chapter is divided into three main sections for convenience; (2.1) bacterial strains and vectors, (2.2) microbiological techniques and standard media and (2.3) nucleic acid isolation and manipulation.

## 2.1 BACTERIAL STRAINS AND VECTORS

The bacterial stains, plasmids and bacteriophages used in this study are listed in Tables 2.1, 2.2 and 2.3 respectively.

TABLE 2.1 Bacterial strains

STRAIN	GENOTYPE	REFERENCE/SOURCE
<i>Escherichia coli</i> strains		
CB51	<i>dam, ara, thi, Δ(lac-pro)</i>	Chris Boyd
DS941	<i>recF143, proA7, str31, thr1, leu6, tsx33, mt12, his4, argE3, lacY<sup>+</sup>, lacZΔM15, lacI<sup>q</sup>, galK2, ara14, supE44, xyl5.</i>	Dave Sherratt
TG1	<i>supE, hsdΔ5, thi, Δ(lac-proAB), F' [traD36, proAB<sup>+</sup>, lacI<sup>q</sup>, lacZΔM15].</i>	Gibson (1984)
<i>Streptomyces lividans</i> strains		
TK24	<i>str6.</i>	Hopwood <i>et al.</i> , (1985a)
TK54	<i>his2, leu2, spc1.</i>	Hopwood <i>et al.</i> , (1985a)
TK64	<i>pro2, str6.</i>	Hopwood <i>et al.</i> , (1985a)
<i>Streptomyces rimosus</i> strains		
M4018	OTC <sup>R</sup>	Rhodes <i>et al.</i> , (1981)
M15883	OTC <sup>R</sup>	Binnie <i>et al.</i> , (1989)

TABLE 2.2 Plasmids

PLASMID	DESCRIPTION*	REFERENCE/SOURCE
pDM130	573bp <i>Sma</i> I <sub>4</sub> - <i>Bam</i> HI <sub>6</sub> fragment from pGLW7 in pIJ486.	Doyle (1987)
pGLW7	2.6kb <i>Sst</i> I <sub>2</sub> - <i>Bam</i> HI <sub>6</sub> fragment encoding <i>otrA</i> in pT7-1.	Iain Hunter
pIJ2843	--	Clayton and Bibb (1990)
pIJ486	--	Ward <i>et al.</i> , (1986)
pKO500	--	Close and Rodriguez (1982)
pPFZ46	6.0kb <i>Sst</i> I <sub>2</sub> - <i>Sst</i> I <sub>13</sub> fragment encoding <i>otcC</i> , <i>otcZ</i> and <i>otrA</i> in pPFZ34.	Pfizer
pPFZ105	5.5kb <i>Pst</i> I <sub>8</sub> - <i>Pst</i> I <sub>18</sub> fragment in pPFZ34.	Pfizer
pPFZ513	1.0kb <i>Bcl</i> I <sub>5</sub> - <i>Pst</i> I <sub>8</sub> fragment in pUC18.	Pfizer
pUC18/19	--	Yanisch-Perron <i>et al.</i> , (1985)
pKME805	305bp <i>Bcl</i> I <sub>5</sub> - <i>Kpn</i> I <sub>7</sub> fragment from pPFZ513 in pUC18. <i>Bcl</i> I <sub>5</sub> site nearest the <i>Eco</i> RI site in the polylinker.	Chapter 3
pKME807	722bp <i>Kpn</i> I <sub>7</sub> - <i>Pst</i> I <sub>8</sub> fragment from pPFZ513 in pUC18.	Chapter 3
pKME808	805bp <i>Bam</i> HI <sub>6</sub> - <i>Pst</i> I <sub>8</sub> fragment from pPFZ513 in pUC18.	Chapter 3
pKME917	1.6kb <i>Sma</i> I <sub>4</sub> - <i>Sma</i> I <sub>9</sub> fragment from pPFZ46 in pUC19. <i>Sma</i> I <sub>4</sub> site nearest the <i>Eco</i> RI site in the polylinker.	Chapter 4
pKMS605	<i>Hind</i> III/ <i>Eco</i> RI insert from pKME805 in pIJ486.	Chapter 3
pKMS607	<i>Hind</i> III/ <i>Eco</i> RI insert from pKME807 in pIJ486.	Chapter 3
pKMS608	<i>Hind</i> III/ <i>Eco</i> RI insert from pKME808 in pIJ486.	Chapter 3
pKMSOCP1	259bp <i>Sau</i> 3AI fragment containing <i>otcCp1/otcXp1</i> promoter region in pIJ2843.	Chapter 6

\* the numbering system used to designate restriction sites in this thesis is new and is not that of Butler *et al.*, (1989)

TABLE 2.3 Bacteriophages

BACTERIOPHAGE	DESCRIPTION	REFERENCE/SOURCE
M13mp18/19	-- Yanisch-Perron <i>et al.</i> , (1985)	
mL6A	2.1kb <i>Sma</i> I <sub>4-9</sub> fragment in M13mp18. <i>Sma</i> I <sub>9</sub> site being nearest the <i>Eco</i> RI site in the polylinker.	Pfizer
mL6C	As mL6A except the <i>Sma</i> I <sub>4</sub> site is nearest the <i>Eco</i> RI site.	Pfizer
mKM804	387bp <i>Hinc</i> II- <i>Bam</i> HI <sub>6</sub> fragment from pGLW7 in M13mp18.	Chapter 3
mKM905	305bp <i>Bcl</i> I <sub>5</sub> - <i>Kpn</i> I <sub>7</sub> fragment from pPFZ513 in M13mp19. <i>Bcl</i> I <sub>5</sub> site nearest the <i>Eco</i> RI site in the polylinker.	Chapter 4
mKM803/903	250bp <i>Pst</i> I <sub>8</sub> - <i>Sma</i> I <sub>9</sub> fragment from pKME917 in M13mp18/19.	Chapter 4
mKM805/905	<i>Hind</i> III/ <i>Eco</i> RI fragment from pKME805 in M13mp18/19.	Chapter 4
mKM807/907	<i>Hind</i> III/ <i>Eco</i> RI fragment from pKME807 in M13mp18/19.	Chapter 4
mKM817/917	<i>Hind</i> III/ <i>Eco</i> RI fragment from pKME917 in M13mp18/19.	Chapter 4
mKM810/910	965bp <i>Sph</i> I <sub>12</sub> - <i>Sst</i> I <sub>13</sub> fragment from pPFZ105 in M13mp18/19.	Chapter 6
mKM88/98	<i>Sau</i> 3AI insert from the pIJ2843 construct in isolate 8, removed as <i>Hind</i> III/ <i>Eco</i> RI fragment and inserted into M13mp18/19.	Chapter 6
mKM812/912	<i>Sau</i> 3AI insert from the pIJ2843 construct in isolate 12, removed as <i>Hind</i> III/ <i>Eco</i> RI fragment and inserted into M13mp18/19.	Chapter 6
mKM815/915	<i>Sau</i> 3AI insert from the pIJ2843 construct in isolate 15, removed as <i>Hind</i> III/ <i>Eco</i> RI fragment and inserted into M13mp18/19.	Chapter 6

## 2.2 MICROBIOLOGICAL TECHNIQUES AND STANDARD MEDIA

Chemicals of good quality were used in the preparation of the growth media and solutions; AnalaR grade when available. The sources of many of the chemicals varied during the course of this work. The most common suppliers were BDH Chemicals Ltd, Poole, Dorset; Difco Laboratories, Detroit, Michigan, USA and Sigma Chemical Co. Ltd, Poole, Dorset.

### 2.2.1 STANDARD MEDIA FOR THE PROPAGATION OF *E. COLI*

**L-Broth:** 10g tryptone, 5g yeast extract, 5g NaCl, 1g glucose, 20mg thymine, made up to 1 litre in distilled water and adjusted to pH 7.0 with NaOH.

**L-Agar:** As L-Broth without glucose and the addition of 15g.L<sup>-1</sup> agar.

**2YT-Broth:** 10g bacto-tryptone, 10g yeast extract, 5g NaCl made up to 1 litre in distilled water.

**Minimal Agar:** 7g K<sub>2</sub>HPO<sub>4</sub>, 2g KH<sub>2</sub>PO<sub>4</sub>, 4g (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.25M trisodium citrate, 0.1g MgSO<sub>4</sub>.7H<sub>2</sub>O, 17.5g agar made up to 1 litre in distilled water.

**Minimal Medium Supplements:** glucose and thiamine were added to give concentrations of 2mg.ml<sup>-1</sup> and 20µg.ml<sup>-1</sup>, respectively.

**Davis and Mingoli (D&M) Salts (X4):** 28g K<sub>2</sub>HPO<sub>4</sub>, 8g KH<sub>2</sub>PO<sub>4</sub>, 16g (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> 1g trisodium citrate, 0.4g MgSO<sub>4</sub>.7H<sub>2</sub>O, made up to 1 litre with distilled water.

### 2.2.2 STANDARD MEDIA FOR THE PROPAGATION OF *STREPTOMYCES*

**Emersons agar:** purchased as a powder and prepared as directed by the manufacturer (Difco).

**Hopwood's Minimal Medium (HMM):** 0.5g L-asparagine, 0.5g K<sub>2</sub>HPO<sub>4</sub>, 0.2g MgSO<sub>4</sub>.7H<sub>2</sub>O, 0.01 FeSO<sub>4</sub>.7H<sub>2</sub>O, 10g glucose, 10g agar, made up to 1 litre with distilled water.

**Liquid Complete Medium (LCM):** 10g glucose, 5g yeast extract, 15g lactalbumin hydrolysates, 2.8g sucrose, 1g calcium chloride, made up to 1 litre with distilled water.

**R2 Medium:** R2A- 44g agar, 0.5g K<sub>2</sub>SO<sub>4</sub>, 20.2g MgCl<sub>2</sub>·6H<sub>2</sub>O, 5.9g CaCl<sub>2</sub>·2H<sub>2</sub>O, 20g glucose, 6g proline, 0.2g casamino acids, 4ml trace elements solution (Hopwood et al., 1985a), made up to 1 litre in distilled water. R2B- 11.5g MOPS, 10g yeast extract, 203g sucrose, adjusted to pH 7.4 with NaOH, made up to 1 litre in distilled water. Equal volumes of R2A (melted and cooled to 55°C) and R2B plus 1ml of 1% (w/v) KH<sub>2</sub>PO<sub>4</sub> were combined prior to use.

**R9 Medium:** R9A- 44g agar, 0.5g K<sub>2</sub>SO<sub>4</sub>, 8.2g MgCl<sub>2</sub>·6H<sub>2</sub>O, 4.7g CaCl<sub>2</sub>·H<sub>2</sub>O, 4g NaNO<sub>3</sub>, 1g KCl, 0.4g MgSO<sub>4</sub>, 20g glucose, 0.8g casamino acids, 4ml trace elements solution (Hopwood et al., 1985a), 2ml FeSO<sub>4</sub> (1% [w/v] solution), made up to 1 litre in distilled water. R9B- 11.5g MOPS, 10g yeast extract, 410g sucrose, adjusted to pH7.4 with NaOH, made up 1 litre with distilled water. Equal volumes of R9A (melted and cooled to 55°C) and R9B were mixed with 1ml of 1% (w/v) KH<sub>2</sub>PO<sub>4</sub> prior to use.

**Soya Mannitol Agar (SM):** 20g mannitol, 20g soya bean flour, 16g agar, made up to 1 litre using tap water.

**Tryptone Soya Broth (TSB):** 30g of Oxoid tryptone soya broth powder made up to 1 litre in distilled water.

**Trusoya Medium 1 (TS1):** soya flour, starch, vegetable oil, inorganic salts and MOPS (pH7.5). The exact composition of this medium cannot be revealed as it is used commercially by Pfizer for strain improvement.

**Yeast extract-Malt extract (YEME):** 3g Difco yeast extract, 5g Difco bacto peptone, 3g Oxoid malt extract, 10g glucose, 340g sucrose, made up to 1 litre in distilled water.



### 2.2.3 STERILISATION

All growth media were sterilised by heating to 120°C for 15 mins in an autoclave. Supplements and buffer solutions were heated to 108°C and CaCl<sub>2</sub> to 114°C for 10 mins. Heat-labile solutions, such as amino acids, were sterilized by filtration through Nalgene 0.22µm pore membranes.

### 2.2.4 PREPARATION OF *STREPTOMYCES* SPORE SUSPENSIONS

Concentrated spore suspensions were required for inoculating liquid cultures of *S. lividans* and *S. rimosus*. The protocol described by Hopwood *et al.* (1985a) was followed with minor modifications.

**Procedure:** *S. lividans* and *S. rimosus* strains used during this work sporulated exceptionally well on Soya Mannitol (SM) agar. A boiling tube containing a slant of SM agar (produced by pouring ca. 15ml of molten agar into the tube and allowing it to solidify with the tube held at 5° from the horizontal) was inoculated with 150µl of a spore or mycelial fragment suspension and incubated at 30°C. After 4-7 days the surface of the culture was covered in a dark grey mass of spores. The slant was then sealed using parafilm and frozen at -20°C. Spores could then be harvested immediately or stored at -20°C indefinitely. The spores were harvested by adding 5ml of 20% (v/v) glycerol to the frozen slant and rubbing the surface of the slant with a 10ml glass pipette. When the surface had been scraped clean of spores, the spore suspension was decanted into a sterile universal. The slant was rinsed with an additional 5ml of 20% (v/v) glycerol to remove any spores remaining from the first wash. The spore suspension (ca. 10ml) thus produced was dark grey in colour. Contaminating agar or mycelial fragments were removed by a single passage through a cotton wool filter, as described in Hopwood *et al.* (1985a). The filtered spore suspension was then aliquoted and frozen at -20°C. Aliquots were thawed and thoroughly mixed using a vortex, prior to use. The titre of the spore suspension was determined after storage at -20°C overnight by spreading out serial dilutions onto plates containing Emersons agar. Titres remained stable over several months. Solid growth media other than SM agar were conducive to sporulation. R2 agar was used

sometimes for *S. lividans* strains and R9 agar was used occasionally for *S. rimosus* strains.

#### 2.2.5 *E. COLI* GROWTH CONDITIONS

Liquid cultures of *E. coli* strains from which plasmids were to be isolated were grown in L broth with the appropriate antibiotic selection (usually ampicillin at  $50\mu\text{g.ml}^{-1}$ ). The volume of broth inoculated depended on the quantity of plasmid required. Routinely, 5ml and 200ml cultures were used for small and large scale plasmid preparations, respectively (see section 2.3.1). For the preparation of competent cells, liquid cultures of *E. coli* DS941 or CB51 were grown in L broth while *E. coli* TG1 was grown in 2YT. To maximise aeration of the culture, the volume of the Ehrlenmeyer flask was at least five times that of the broth. All cultures were incubated at  $37^{\circ}\text{C}$  in an orbital shaker at ca. 250 rpm. *E. coli* strains were also propagated on L-agar or Minimal Medium agar plates (containing supplements) with the appropriate selection. Plates contained ca. 25ml of agar and were incubated overnight at  $37^{\circ}\text{C}$ .

#### 2.2.6 GROWTH OF *STREPTOMYCES* MYCELIUM

Ehrlenmeyer flasks were used for growing liquid cultures of *Streptomyces*. YEME broth was used for growing *Streptomyces lividans* strains and TSB broth used for growing *Streptomyces rimosus* strains. Cultures required for formation of *S. lividans* or *S. rimosus* protoplasts also contained 0.5% (w/v) glycine and 5mM  $\text{MgCl}_2$ . The volume of the culture depended on the final use of the mycelium. The formation of protoplasts required 30 to 50ml of broth, while 200ml of broth or multiples thereof were used for plasmid isolation. Routinely, 100 $\mu\text{l}$  of a dense spore suspension ( $10^5$ – $10^6$  spores) were used to inoculate 50ml of broth. The flask volume was at least five times the volume of the broth to facilitate good aeration when incubated in an orbital shaker ( $30^{\circ}\text{C}$  at ca. 250 rpm).

#### 2.2.7 QUANTIFICATION OF *STREPTOMYCES* GROWTH

Optical density measurements gave a poor estimation of growth due to the formation of mycelial pellets and wefts in liquid cultures of

*Streptomyces*. Dry weights were determined, therefore, according to the method of Harvey (pers. comm.). Duplicate Whatman Glass Fibre Filters (GF/C, diameter 4.7cm) were numbered with a pencil and dried for 15 min in a microwave oven (600 W), which was set on reheat, and allowed to cool to room temperature in a desiccator for 30 min. The weights of the filters were recorded to three decimal places. Culture samples, usually 5ml, were filtered through the GF/C filters using a Millipore sintered glass vacuum filtration unit. The filters were then washed by passing 15ml of dH<sub>2</sub>O through the filtration system, dried in the microwave, allowed to cool in the desiccator and weighed as before. The dry weight (g.L<sup>-1</sup>) was determined by deducting the original weight (g) of the filter from the weight recorded after filtration and then multiplying this value by the dilution factor. At least three independent dry weight determinations were used to calculate the average value for a particular time point.

#### 2.2.8 QUANTIFICATION OF OXYTETRACYCLINE IN CULTURE SUPERNATANTS

Samples of liquid cultures (1ml) were extracted into 9ml of HCl (pH 1.7) for 30min and then filtered through Whatman No.1 papers. 1ml samples were then sent to the Pfizer Process Development Laboratories at Sandwich and quantified by isocratic, ion-paired, reverse-phase high performance liquid chromatography (hplc) using a C<sub>18</sub>  $\mu$  Bondabank column (Waters, Millipore House, Harrow, Middlesex, UK.) with a mobile phase of acetonitrile/water (3:7) containing 0.5g.L<sup>-1</sup> of 1-hexanesulphonic acid, adjusted to pH 1.7 with sulphuric acid.

#### 2.2.9 ANTIBIOTICS AND INDICATORS

The antibiotic concentrations used throughout for both liquid and plate selection were as follows:

DRUG	SELECTIVE CONCENTRATION	STOCK SOLUTION	STORAGE TEMP
Ampicillin	50 $\mu$ g.ml <sup>-1</sup>	20mg.ml <sup>-1</sup> (water)	-20°C
Kanamycin	5-900 $\mu$ g.ml <sup>-1</sup>	20mg.ml <sup>-1</sup> (water)	-20°C
Thiostrepton	25 $\mu$ g.ml <sup>-1</sup>	10mg.ml <sup>-1</sup> (DMSO)	4°C

Stock solutions were added to molten agar, which had been cooled to 55°C.

X-gal (5-bromo-4-chloro-3-indolyl- $\beta$ -galactosidase) was used in conjunction with IPTG to identify *E. coli* strains CB51, DS941 and TG1 containing pUC or M13mp vectors with inserts in their multiple cloning sites. Recombinants containing inserts are generally white while those lacking inserts are blue. X-gal was stored at a concentration of 20mg.ml<sup>-1</sup> in dimethylformamide (DMF) at -20°C while IPTG was stored at a concentration of 24mg.ml<sup>-1</sup> in dH<sub>2</sub>O at -20°C. X-gal and IPTG were added to L-agar plates to a final concentration of 20 $\mu$ g.ml<sup>-1</sup> and 50 $\mu$ g.ml<sup>-1</sup>, respectively.

#### 2.2.10 PRESERVATION OF *STREPTOMYCES* AND *E. COLI* STRAINS

*Streptomyces* spp. were preserved by freezing agar slants covered in spores at -20°C. The *Streptomyces* spp. used in this work remain viable indefinitely under these conditions. Small slants made in 5ml bijoux were used for the long-term storage of strains. However, few spores could be harvested from these slants and it was therefore necessary to inoculate large slants (ca. 15ml) in order to generate sufficient spores for most purposes. *E. coli* strains were stored in glycerol, an 800 $\mu$ l aliquot of an overnight culture was mixed with an equal volume of 40% (v/v) glycerol, 2% peptone (w/v) and frozen at -70°C. The strains were revived by scraping the surface of the frozen suspension with a toothpick and either inoculating liquid broth or streaking onto agar to a isolate single colony.

#### 2.2.11 INTRODUCTION OF PLASMID DNA INTO *E. COLI*

Plasmids were introduced into *E. coli* DS941 and CB51 by genetic transformation. CaCl<sub>2</sub> treatment routinely yielded cells with a transformation efficiency of 10<sup>6</sup>-10<sup>7</sup>. $\mu$ g<sup>-1</sup> plasmid.

**Preparation of competent cells:** An overnight culture of the recipient strain was diluted 1 in 100 into 30ml L-broth and incubated for 90-120 min to a density of approximately  $10^8 \text{ ml}^{-1}$  cells ( $\text{OD}_{600}$  0.45-0.55). The cells were harvested using a centrifuge (12000g, 5 min, 4°C) and resuspended in 10ml of ice-cold 50mM  $\text{CaCl}_2$ . The cells were pelleted again, resuspended in 1ml of ice-cold 50mM  $\text{CaCl}_2$  and kept on ice for at least 15 min before use. At this stage glycerol could be added to the cells to 20% (v/v) and 200 $\mu\text{l}$  aliquots frozen at -70°C. The cells retained their competence over several months without any significant decrease in their transformation efficiency.

**Transformation with plasmid DNA:** 200 $\mu\text{l}$  aliquots of the competent cells were added to plasmid DNA (1-100ng) in a volume less than 10 $\mu\text{l}$ , mixed gently by inverting the microfuge tube and incubated on ice for up to 1 hour. The cells were heat-shocked (2 min, 42°C) and returned to ice for a further 15 min. An 800 $\mu\text{l}$  aliquot of L-broth was added and the cell suspension incubated at 37°C for 45 min to allow expression of the plasmid resistance genes. For ampicillin selection an expression stage as such was not required. The cells were spread on L-agar plates containing the appropriate antibiotic.

#### 2.2.12 TRANSFECTION OF *E. COLI* TG1 WITH BACTERIOPHAGE M13

A modified version of the protocol developed by Hanahan (1983) was used to prepare competent cells of *E. coli* TG1 for transfection with the replicative form of M13.

**Reagents:** TFB; 10mM 2[N-morpholino]ethanesulphonic acid (MES) buffer (pH 6.3), 10mM  $\text{RbCl}$ , 45mM  $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$ , 10mM  $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$ , DTT/KAc; 2.25mM dithiothreitol, 40 mM potassium acetate (pH 6.0).

**Preparation of competent cells:** An overnight culture of *E. coli* TG1 grown in 2YT broth was diluted 1 in 100 into 30 ml of fresh 2YT. The culture was grown until the  $\text{OD}_{600}$  was between 0.45 and 0.55. The culture was then transferred to a 50ml polypropylene centrifuge tube

(pre-rinsed with sterile distilled dH<sub>2</sub>O and pre-chilled on ice) and incubated on ice for 20 min. The cells were harvested by centrifugation (4000g, 4°C for 10 min), resuspended very gently in 2.5ml of pre-chilled TFB and incubated on ice for 15min. The cells were then treated with 100µl of dimethylsulphoxide (DMSO) and incubated for 5min on ice. This was followed by the addition of 100µl DTT/KAc and incubation on ice for a further 5min. The cells were ready to use after adding a second 100µl aliquot of DMSO and incubating on ice for 5min.

**Transfection of *E. coli* TG1 with bacteriophage M13mp:** The procedure followed the plasmid transformation protocol described above, up to and including the heat-shock stage. After this step, 200µl of a fresh exponential TG1 culture were added to the transfected cells, followed by 10µl of IPTG (24mg.ml<sup>-1</sup>) and 50µl X-gal (20mg.ml<sup>-1</sup>). The cells were then mixed and added to 2.5ml of molten water-agar (0.6% w/v, pre-cooled to 45°C), which were poured onto thoroughly dried minimal medium plates containing D & M salts, vitamin B1 (5µg.ml<sup>-1</sup>) and glucose (2mg.ml<sup>-1</sup>) supplements.

### 2.2.13 INTRODUCTION OF PLASMID DNA INTO *STREPTOMYCES*

Plasmids were introduced into *Streptomyces* spp. by genetic transformation using the polyethylene-glycol-mediated protocol described by Hunter (1985).

**Reagents:** Trace element solution; Per litre, 40mg ZnCl<sub>2</sub>, 200mg FeCl<sub>3</sub>.6H<sub>2</sub>O, 10mg CuCl<sub>2</sub>.2H<sub>2</sub>O, 10mg MnCl<sub>2</sub>.4H<sub>2</sub>O, 10mg Na<sub>2</sub>B<sub>4</sub>O<sub>7</sub>.10H<sub>2</sub>O and 10mg (NH<sub>4</sub>)<sub>6</sub>Mo<sub>7</sub>O<sub>24</sub>.4H<sub>2</sub>O. **Medium P;** 5.73g N-tris[Hydroxymethyl]methyl-2-aminoethanesulfonic acid (TES), 103g sucrose, 2.93g MgCl<sub>2</sub>.7H<sub>2</sub>O, 0.5g K<sub>2</sub>SO<sub>4</sub>, 3.68g CaCl<sub>2</sub>.2H<sub>2</sub>O, 2ml trace element solution. Adjust to pH 7.4 with NaOH and made up to 1 litre in distilled water. **Lysozyme solution;** 10% (w/v) sucrose, 25mM TES buffer (pH 7.2), 2.5mM K<sub>2</sub>SO<sub>4</sub>, 2ml trace elements (Hopwood et al., 1985a), 2.5mM MgCl<sub>2</sub>, 2.5mM CaCl<sub>2</sub>. KH<sub>2</sub>PO<sub>4</sub> (0.005% [w/v]) and lysozyme (0.3mg.ml<sup>-1</sup>) were added immediately prior to use. **PEG solution;** 1g of polyethylene glycol 1540 (supplied by BDH) was melted in a microwave (600W) for 10s on the reheat setting and then mixed with 3ml of medium P.

**Preparation of protoplasts:** 30ml cultures were grown in the appropriate medium at 30°C. The optimal time to harvest the mycelium in order to recover the most competent protoplasts was species-dependent i.e. *S. lividans* mycelium were harvested after 65hr and *S. rimosus* after 48hr, respectively. The mycelium were pelleted at 12000g for 10 minutes and washed twice in 10.3% (w/v) sucrose. The pellet then resuspended in 4ml of lysozyme solution and incubated at 37°C for 15-30 min. The formation of protoplasts was monitored using a microscope, the reaction terminated by adding 5ml of P medium and the protoplasts triturated twice. The protoplasts were then filtered through cotton wool (Hopwood et al., 1985a), pelleted using a centrifuge (12000g for 10 minutes) and washed twice in P medium. Finally, they were resuspended in 4ml of medium P, dispensed into 200µl aliquots and frozen at -70°C.

**Transformation of protoplasts:** The protoplasts were thawed on ice. DNA was added in a volume of less than 10µl and the mixture incubated on ice for 30s. PEG solution (400µl) was added, the solution incubated for a further 1min on ice and finally, medium P (800µl) was added. Dilutions of the transformation mix were then made in medium P.

**Regeneration of transformed protoplasts:** The method used to prepare the regeneration medium for the protoplasts was standardised. The medium was stored in two parts, RA and RB ; the former solid and the latter liquid. Both parts were placed in a steam bath until the RA portion melted. The two components were allowed to cool to 50°C before they were combined and 1ml of 1%(w/v)KH<sub>2</sub>PO<sub>4</sub> added. The complete regeneration medium was mixed by swirling and then poured into petri dishes (diameter 9cm). On average 8 plates were obtained from 200ml of medium. The plates were dried in a laminar flow hood to minimise air borne contamination by leaving the lids half open for 45min. They were then rotated 180° and their relative positions reversed so that those at the front of the hood were placed at the back. After a further 45min, the plates were removed and left overnight at 30°C. The next day, any plates which showed any signs of contaminating growth were discarded and the rest used for the regeneration of protoplasts.

The only drug resistance used in this work for plasmid selection in *Streptomyces* was thiostrepton (obtained from E.R. Squibb, New Jersey, USA). It was dissolved in DMSO to make a 1% (w/v) stock solution. Transformed protoplasts were selected after 16-22 hrs of non-selective growth at 30°C by overlaying the regeneration plates with 1ml of a 220µg.ml<sup>-1</sup> thiostrepton solution in 10.3% (w/v) sucrose solution.

*S. lividans* protoplasts were regenerated on R2 agar plates and R9 plates were used for the regeneration of *S. rimosus*.

## 2.3 NUCLEIC ACID ISOLATION AND MANIPULATION

### 2.3.1 ISOLATION OF PLASMID DNA FROM *E. COLI* AND *STREPTOMYCES*

Protocols based on the alkaline lysis method (Birnboim and Doly, 1979) were used for the isolation of plasmid from small (5ml) or large (50-200ml) culture volumes of both *Streptomyces* spp and *E. coli*.

Reagents: Birnboim Doly I (BDI); 50mM glucose, 25mM Tris-HCl (pH 8.0), 10mM EDTA. Lysozyme was added immediately before use to a final concentration of 1-4mg.ml<sup>-1</sup>. Birnboim Doly II (BDII); 0.2M NaOH, 1% (w/v) SDS which was stored in a plastic container. Birnboim Doly III (BDIII); 5M KOAc (pH 4.8); prepared by mixing equal volumes of 3M CH<sub>3</sub>COOK and 2M CH<sub>3</sub>COOH. Phenol; The phenol was redistilled, buffered against 0.5M Tris.HCl (pH 8.0) and contained 0.1% (w/v) 8-hydroxyquinoline. Phenol.Chloroform; 50 parts phenol, 49 parts chloroform, 1 part isoamyl alcohol. DNAase-free RNAase; Pancreatic RNAase (RNAase A) was dissolved at a concentration of 10mg.ml<sup>-1</sup> in dH<sub>2</sub>O, heated to 100°C for 15min and allowed to cool slowly to room temperature. The RNAase was then aliquoted and stored at -20°C.

Large scale plasmid preparations: 200ml cultures of stationary phase cells were harvested using a centrifuge (12000g, 5min at 4°C). The pellet was resuspended in 4ml of Birnboim-Doly I solution. At this stage, *Streptomyces* preparations were treated with lysozyme. The amount of lysozyme that resulted in the optimum recovery of plasmid



was determined for each strain. *S. lividans* and *S. rimosus* strains were incubated at 37°C for 20min with 4mg.ml<sup>-1</sup> and 1mg.ml<sup>-1</sup> lysozyme solutions, respectively. *E. coli* cultures were incubated on ice for 5min. Then 8ml of Birnboim-Doly II solution were added and the solution left on ice for 5-10 min before 6ml of cold Birnboim-Doly III solution were added. The suspension was mixed gently and left on ice for 15-30 min. The cell debris and most of the chromosomal DNA was removed by centrifugation (32000g, 5 min at 4°C). The remaining nucleic acid was precipitated by the addition of an equal volume of isopropanol and then harvested by centrifugation (39200g, 15 min). The nucleic acid pellet was washed with 70% (v/v) ethanol. The plasmid DNA was further purified by equilibrium density centrifugation on a caesium chloride-ethidium bromide (CsCl-EtBr) gradient. The nucleic acid pellet was redissolved in 1 ml of dH<sub>2</sub>O and 4.5g of CsCl dissolved in 3.5ml of dH<sub>2</sub>O. The DNA and CsCl solutions were combined with 250µl of EtBr (10mg.ml<sup>-1</sup>), creating a solution with a density of 1.58g.ml<sup>-1</sup>. The nucleic acid-CsCl solution was spun in a Beckman Ti70 angled rotor at 289,000g for 16 hours at 20°C. Two bands were visible in the gradients after centrifugation, a lower supercoiled plasmid band and an upper chromosomal and relaxed plasmid DNA band. The lower band was removed using a 1ml syringe and the EtBr removed by repeated extractions with water-saturated butanol. After dilution with 3 volumes of dH<sub>2</sub>O, 9 volumes of absolute ethanol were added. The precipitate was pelleted by centrifugation (27000g, 4°C for 30min). The resulting plasmid pellet was washed twice with 70% (v/v) ethanol and dried *in vacuo* before being redissolved in 1ml dH<sub>2</sub>O. This procedure yielded very large amounts of pure plasmid DNA (up to 1mg from *E. coli* and 100µg from *Streptomyces* cultures) suitable for all *in vitro* manipulations.

**Small scale plasmid preparations:** This protocol was used for isolating plasmids from both *E. coli* and *Streptomyces* spp, without modification. Routinely, plasmids were isolated from 1.5ml of *E. coli* cultures and 3.0ml of *Streptomyces* cultures or 2cm<sup>2</sup> patches of mycelial growth on agar plates. The cells were pelleted by centrifugation in a 1.5ml microfuge tube (12000g for 30s) and resuspended in 100µl of BDI, containing lysozyme at a concentration of 1mg.ml<sup>-1</sup>, using a vortex

mixer. This was followed by the addition of 200 $\mu$ l of BDII and repeated inversion of the microfuge tube to mix thoroughly the suspension. Immediately afterwards, 150 $\mu$ l of prechilled BDIII was added to the viscous bacterial lysate, mixed gently on the vortex mixer and placed on ice for 5-10 min. The cell debris and most of the chromosomal material was harvested by centrifugation (12000g, 4°C for 10min) in a microfuge. The supernatant was transferred to a fresh tube and extracted with half volumes of phenol:chloroform and chloroform. The nucleic acid was then precipitated by the addition of 2 volumes of ethanol and allowed to stand at room temperature for 3-5min. The precipitate was harvested by centrifugation in a microfuge (12000g, 4°C for at least 15min). The resulting pellet was rinsed twice with 70% (v/v) ethanol before it was allowed to dry by leaving the tube open on the bench. The nucleic acid was then resuspended in 50 $\mu$ l dH<sub>2</sub>O containing DNAase-free RNAase (20 $\mu$ g.ml<sup>-1</sup>). The typical yield of high-copy-number plasmids such as pUC from *E. coli* and pIJ486 from *Streptomyces* was 2-5 $\mu$ g of plasmid. Plasmid prepared in this way could be used for most *in vitro* manipulations.

### 2.3.2 ISOLATION OF TOTAL RNA

This procedure is based on that of Kirby *et al.* (1976) with extensive modifications by Covey and Smith ( see Hopwood *et al.*, 1985a). RNAase is a very persistent enzyme and precautions were taken against contamination of equipment and solutions. Prior to use, all glassware was incubated overnight at 300°C. Distilled water was treated with DEPC (0.1% [v/v] of total volume) overnight and heated to 121°C for 15min in an autoclave. All solutions were prepared from DEPC-treated water and chemicals reserved for RNA work.

**Reagents:** Phenol; The phenol was redistilled, buffered against 0.5M Tris.HCl (pH 8.0) and contained 0.1% (w/v) 8-hydroxyquinoline. Phenol:Chloroform; 50 parts phenol, 49 parts chloroform, 1 part isoamyl alcohol. 10X DNAase RQ buffer; 400mM Tris.HCl ( pH 7.9), 100mM NaCl, 60mM MgCl<sub>2</sub>. Kirby mixture; 1g tri-isopropyl-naphthalene sulphonate, 6g 4-amino salicylate (Na salt), 50mM Tris-HCl (pH 8.3), 6ml phenol, made up to 100ml in distilled water.

**Procedure:** RNA was isolated always from 50ml liquid cultures grown in 250ml Ehrlenmeyer flasks. When the culture reached the appropriate stage of growth, the contents were immediately decanted into a 50ml flask containing 20ml of dH<sub>2</sub>O, which had been frozen at -20°C, rapidly lowering the temperature to ca. 0°C. As much of the chilled culture as possible, including any residual ice, was then transferred to an unused polypropylene centrifuge tube. The mycelium were then harvested using a centrifuge in a prechilled rotor. The pellet was resuspended in 5ml of Kirby mix (prechilled on ice) containing 12g of 0.45mm diameter glass beads and agitated thoroughly on a vortex mixer for 1-2min. 5ml of phenol:chloroform (stored at room temperature) were then added and the mixture agitated for a further 1-2min. The homogenate was then spun in the centrifuge (9000g, 4°C for 10min) to separate the phases. The aqueous phase was removed, using a baked Pasteur pipette, to a fresh centrifuge tube containing 5ml of phenol:chloroform and mixed for 1min on the vortex mixer. The phases were then separated as before. At this stage, the nucleic acid solution could be stored at -70°C without any degradation being detectable. To remove DNA from the preparation, the RNA could be pelleted by caesium chloride centrifugation and/ or it could be treated with RNAase-free DNAase.

**Pelleting the RNA through a dense caesium chloride cushion:** Homogenate (ca. 3.5ml) from the Kirby lysis was layered carefully on top of a 5.7M CsCl-0.1 EDTA cushion (1.5ml, density 1.707 g.ml<sup>-1</sup>) in a ultracentrifuge tube (5ml, Beckman, Polyallomer™). The RNA was then pelleted by centrifugation (35,000rpm, 20°C for 16 hours) using a Beckman SW50.1 rotor. After centrifugation, the supernatant was removed using a baked pasteur pipette. The tube was then inverted to drain away the remaining fluid. Using a fresh scalpel blade to cut the centrifuge tube, the bottom was isolated in the form of a small cup. The pellet was dissolved in 400µl of dH<sub>2</sub>O and extracted with half volumes of phenol:chloroform and chloroform. The RNA was precipitated by the addition of 1/50 volume of 5M NaCl and two volumes of absolute ethanol,

and harvested by centrifugation in a microfuge (12,000g 4°C for 10min). The RNA pellet was washed twice with 70% (v/v) ethanol, dried in the open microfuge tube at room temperature and dissolved in 100-200µl of dH<sub>2</sub>O. The RNA concentration was estimated spectrophotometrically, reprecipitated and stored at -70°C.

**Treatment with DNAase RQ (Promega):** An aliquot of the homogenate from the Kirby lysis, usually 1ml, was precipitated by the addition of 1/50 volume of 5M NaCl and 2 volumes of ethanol. The nucleic acid was harvested by centrifugation, washed twice with 70% (v/v) ethanol and dried by leaving the tube open on the bench. The nucleic acid was then redissolved in 400µl of 1X DNAase RQ buffer, DNAase added to 100U.ml<sup>-1</sup> and the reaction mixture incubated at 37°C for 20-30min. The mixture was then extracted with equal volumes of phenol:chloroform and chloroform. The remaining nucleic acid, in the aqueous phase, was precipitated by addition of 1/10 volume of 5M NaCl and an equal volume of isopropanol. After centrifugation, the pellet was washed in 70% (v/v) EtOH, dried briefly and redissolved in DEPC-treated dH<sub>2</sub>O.

### 2.3.3 QUANTIFICATION OF NUCLEIC ACID

The concentration and purity of the nucleic acid was determined by spectrophotometry; an A<sub>260</sub> of 1 is equivalent to 40µg.ml<sup>-1</sup> RNA and 50µg.ml<sup>-1</sup> double stranded DNA. Pure preparations of RNA and DNA have an A<sub>260</sub>/A<sub>280</sub> of 1.8 and 2.0, respectively. Contaminating protein or phenol lowers significantly these values.

For quantitative S1 nuclease protection experiments, 25-30µg of RNA that had been purified by CsCl density centrifugation and treated with RNAase-free DNAase was resuspended in DEPC-treated dH<sub>2</sub>O. Optical density measurements were used to quantify the amount of RNA in 5µl aliquots which had been diluted to 1ml. The measurements were repeated until three consecutive absorbance values agreed to within ±0.002 (approximately 0.1µg).

### 2.3.4 PRECIPITATION OF NUCLEIC ACID USING ETHANOL OR ISOPROPANOL

DNA solutions were precipitated by the addition of 1/50 volume of 5M NaCl and 2 volumes of cold ethanol or an equal volume of isopropanol. After mixing, the DNA was pelleted by centrifugation (27000g, 30 min, 4°C for volumes of 7.5-20ml or 12000g, 15 min, 4°C for small volumes in microfuge tubes). The pellet was washed in 70% (v/v) ethanol and dried briefly in a vacuum desiccator or in an open tube on the bench.

### 2.3.5 DIGESTION OF DNA WITH RESTRICTION ENDONUCLEASES

Digestion of DNA was performed usually in a total volume of 20µl containing 0.25-1.0µg of DNA and 1-10 units of restriction endonuclease µg<sup>-1</sup> DNA in the appropriate buffer (provided by BRL the main enzyme supplier). For the digestion of larger amounts of DNA, the volumes were scaled up accordingly. The reactions were allowed to proceed for 1 to 2 hours at the appropriate temperature. Reactions were stopped either by the addition of gel loading buffer or by rapid heating to 70°C for 5 min, followed by cooling on ice.

### 2.3.6 LIGATION OF DNA FRAGMENTS

The ligation of DNA fragments was carried out usually at a DNA concentration of 6mg.ml<sup>-1</sup>. The molar ratio of insert fragment to vector was 2:1, when the vector could not ligate to itself, but only with the insert fragment. A molar ratio of 10:1 was used when the ends of the vector could ligate to each other. Ligations were performed usually in 20µl of 1 X BRL ligation buffer, containing 1U of T<sub>4</sub> ligase µg<sup>-1</sup> of DNA. The reactions were incubated for 4 hours at room temperature or overnight at 16°C.

### 2.3.7 REMOVAL OF THE 5' PHOSPHATE FROM LINEARISED DNA

**10X CiP Buffer:** 200mM Tris.HCl (pH 8.0), 10mM MgCl<sub>2</sub>, 10mM ZnCl<sub>2</sub> and 0.5mg.ml<sup>-1</sup> Bovine Serum Albumin.

**Procedure:** Calf Intestinal Alkaline phosphate (CIP) was used to remove the 5' phosphate from DNA. Around 5 pmoles of 5'-terminal phosphorylated DNA with 5' protruding ends (approximately 7µg of a

5kb molecule) were incubated in 1X CIP buffer, containing 0.1U of CIP at 37°C for 30min. The reaction was terminated by heating to 65°C in 1X gel loading buffer for 10min. The 5'-terminal dephosphorylated DNA was recovered from an agarose gel after electrophoresis.

### 2.3.8 REMOVAL OF PROTEIN FROM NUCLEIC ACID SOLUTIONS USING ORGANIC SOLVENTS

**Reagents:** Phenol; All phenol used in the purification of DNA or RNA was redistilled, buffered against 500mM Tris.HCl (pH 8.0) and contained 0.1% (w/v) 8-hydroxyquinoline. Phenol.Chloroform; 50 parts phenol, 49 parts chloroform, 1 part isoamyl alcohol.

**Procedure:** Proteins were removed from nucleic acid solutions by extracting first with phenol.chloroform and then chloroform. 1/2 volumes of the solvents were added to the nucleic acid solution, mixed using a vortex mixer and the phases separated by centrifugation (1200g for 3min). Proteins were retained in the organic phase. Subsequently, precipitation with isopropanol or ethanol removed any solvent in the aqueous phase.

### 2.3.9 PREPARATION OF RADIO-LABELLED PROBES

#### 2.3.9.1 RANDOM PRIMED DNA LABELLING METHOD

Labelling of DNA fragments with  $^{32}\text{P}$  followed the procedure of Feinberg and Vogelstein (1983 and 1984) using a Boehringer Mannheim Multiprime™ kit. The labelling reaction was set up in the following way:

25-50ng of denatured DNA fragment in a volume less than 10 $\mu\text{l}$   
(denatured by heating for 10min at 95°C with subsequent cooling on ice).

1 $\mu\text{l}$  of each unlabelled dNTP, from 0.5mM stocks.

2 $\mu\text{l}$  of reaction mixture (containing hexanucleotide mix and 10X concentrated reaction buffer).

5 $\mu\text{l}$  (50 $\mu\text{Ci}$ ) of [ $\alpha$ - $^{32}\text{P}$ ] dCTP (3000 Ci.mmol $^{-1}$ ).

1 $\mu\text{l}$  (2U) of Klenow enzyme.

Made up to 20 $\mu\text{l}$  total volume with dH $_2$ O.

The reaction mixture was incubated at 37°C for 30 minutes, and stopped by heating to 65°C for 10 minutes. The labelled DNA fragments (and template DNA) were purified from the unincorporated dNTP's by Sephadex-G50 column chromatography.

### 2.3.9.2 5'-END LABELLING OF OLIGONUCLEOTIDES

Oligonucleotide probes were labelled at their 5' terminal ends using T4 polynucleotide kinase and [ $\gamma$ -<sup>32</sup>P]ATP.

**Reagents:** 10 X Kinase Buffer; 500mM Tris-HCl (pH 7.5), 100mM MgCl<sub>2</sub>, 50mM dithiothreitol, 1mM spermidine, 1mM EDTA. Stored at -20°C.

**Reaction mixture:**

	Volume. ( $\mu$ l)
10X Kinase buffer	1
50ng of oligonucleotide	3
50 $\mu$ Ci of [ $\gamma$ - <sup>32</sup> P] ATP (3000Ci.mmol <sup>-1</sup> )	5
T4 Polynucleotide kinase (10U)	1

The reaction was incubated at 37°C for 30min. The unincorporated ATP was removed by Sephadex G50 column chromatography.

### 2.3.9.3 SEPHADEX G50 COLUMN CHROMATOGRAPHY

After the labelling reactions had been completed, 5 $\mu$ l of dextran blue (50mg.ml<sup>-1</sup>) and 5 $\mu$ l of phenol red (50mg.ml<sup>-1</sup>) were added to the reaction mixture. The samples were then loaded onto Sephadex-G50 (20 x 0.5cm dimension gravity column), which had been pre-equilibrated with column buffer (100mM NaCl, 10mM Tris.HCl (pH 7.5) 1mM EDTA). Fractions of approximately 500 $\mu$ l were collected. The dextran blue co-eluted with the DNA fragments and all other aliquots were discarded. The labelled DNA sample was boiled for 5min prior to use.

### 2.3.10 AGAROSE GEL ELECTROPHORESIS

Both DNA and RNA were visualized on horizontal neutral agarose gels. Although 0.8% (w/v) gels were most commonly used, 1-2% (w/v) gels were occasionally used to separate fragments of <1.5kb. Gels were

routinely prepared and run in TBE buffer, however, TAE buffer was used when DNA fragments were to be isolated from the gels (see section 2.3.12).

**Buffers:** 10X TBE Buffer (pH 8.3); 109g Tris, 55g boric acid, 9.3g Na<sub>2</sub>EDTA.2H<sub>2</sub>O made up to 1 litre in distilled water, pH should be ca. 8.3. 10X TAE Buffer pH 8.2; 48.4g Tris, 16.4g Na acetate, 3.6g Na<sub>2</sub>EDTA.2H<sub>2</sub>O, made up to 1 litre in distilled water, pH adjusted to 8.2 with glacial acetic acid. 5X Agarose gel loading (AGL) buffer (pH 7.4); 0.025% (w/v) bromophenol blue, 0.025% (w/v) xylene cyanol, 25% (w/v) ficoll, 0.5% (w/v) SDS, 50mM EDTA.

#### 2.3.10.1 MINI GELS

BRL model H6 gel kits were used for the rapid analysis of DNA after digestion with restriction enzymes or precipitation steps. 0.16g agarose was added to 20ml of 1X TBE (TAE), boiled then cooled to 60°C. EtBr was added to 200ng.ml<sup>-1</sup> and the molten agarose poured into a 7.6cm X 5.1cm gel caster with an 8 well slot former (4.1 X 0.8mm wells). After the gel had set, the slot former was removed and the gel placed in the tank with 500ml of 1X TBE (TAE). Depending on the time available and the level of resolution required the DNA was separated by electrophoresis for 30-45min with an applied voltage of 2-10V.cm<sup>-1</sup>. The separated DNA molecules were visualised on a 302nm UV transilluminator.

#### 2.3.10.2 LARGE GELS

200ml gels were also used to ensure good separation of DNA fragments for accurate sizing and/or Southern analysis. They were made by pouring 200ml of molten agar containing 200µg EtBr, into a 16.5 X 23cm gel former with a 20 space slot former. The gels were run overnight at 20V in 1X TAE or TBE buffer in gel tanks with a capacity of 3 litres. DNA samples were mixed with 1/5 volume of 5X AGL buffer, heated to 70°C for 2min and cooled on ice before loading onto the gel.



**2.3.10.3 PHOTOGRAPHY OF RESOLVED NUCLEIC ACIDS**

Gels stained with ethidium bromide were viewed on a 302nm UV transilluminator and photographed using Polaroid type 67 land film or using a Pentax 35mm SLR loaded with Ilford HP5 film. Both cameras were fitted with Kodak Wratten filters (No. 23A).

**2.3.11 DENATURING POLYACRYLAMIDE GEL ELECTROPHORESIS**

A BRL sequencing unit (Model S2) was used for high voltage polyacrylamide gel electrophoresis.

Reagents: 5X Sequencing gel loading buffer (pH 8.2); 0.1% (w/v) bromophenol blue, 0.1% (w/v) xylene cyanol, 10mM Na<sub>2</sub>EDTA, 95% (v/v) formamide (de-ionized with a mixed-bed resin). 40% Acrylamide stock solution; 76g acrylamide (BDH. Electran Grade) and 4g bisacrylamide. Dissolved in dH<sub>2</sub>O with stirring, brought to a final volume of 200ml, filtered through 3MM paper and stored at 4°C.

Preparation of polyacrylamide gels: 6% (w/v) denaturing polyacrylamide gels were used for sequencing and single-stranded probe isolation. The gels were prepared from the following stock solutions:

40% (w/v) acrylamide stock	15ml
urea	55g
10X TBE	10ml
dH <sub>2</sub> O	<u>35ml</u>
	ca. 100ml.

The urea was dissolved by heating the mix to 37°C and then cooled to room temperature. The sequencing gel solution could be stored at 4°C for several weeks without loss of resolution. Before pouring the gel, 300µl of freshly prepared 10% (w/v) Ammonium persulphate and 50µl of TEMED were added to 50ml of the stock solution.

Preparation of glass plates and pouring the gel: The plates (40cm X 33cm) were cleaned thoroughly with alcohol and water and assembled using three spacers (0.4mm thick) along the vertical sides and the

bottom of the gel. The entire assembly was held in place by clamps. The gel solution was poured from a beaker down one edge of the plates while tilting the plates at an angle of about 30°. The plates were then laid at an angle of 5° and the sharks tooth combs inserted. The gel polymerized usually within 30min at room temperature.

**Electrophoresis of sequencing gels:** The gel was pre-electrophoresed for 30min at a constant power of 60W. Prior to loading, the samples containing sequencing loading buffer were heated to 95°C for 5 mins, placed on ice and loaded on to the gel. 6% (w/v) gels were run for 1.75-2hrs to read the first 100 nucleotides and for 4.5-5hrs to read up to 400 nucleotides.

### 2.3.12 RECOVERY OF DNA FROM AGAROSE GELS

DNA fragments were recovered from TAE agarose gels using GeneClean™ and electro-elution. The GeneClean™ kit (purchased from Stratech Scientific Ltd) was used for recovering DNA in the size range 1-7kb. Smaller fragments were not recovered efficiently from the glass beads and were isolated, therefore, by electro-elution. TBE gels were avoided as the borate ion significantly reduces the amount of DNA recovered using GeneClean™ and can influence some enzyme reactions such as ligation.

#### 2.3.12.1 GENE CLEAN™

The TAE agarose gel was placed on a 302nm UV transilluminator and a small block (<0.2cm<sup>3</sup>) of agarose containing the DNA fragments of interest excised using a scalpel. The agarose block was fragmented and 2.5 volumes of NaI solution were added. The suspension was incubated at 55°C for 5 minutes or until the agarose had dissolved completely. 5µl of "glassmilk" was added to the solution, which was mixed rapidly on a vortex mixer and placed on ice for 5min. The glass beads with the DNA bound were pelleted by centrifugation (5 seconds in a microfuge) and the supernatant discarded. The glass beads were washed three times in ice-cold NEW solution (500µl), each time agitating the suspension using a vortex mixer and recovering the beads by centrifugation. After the final wash, care was taken to remove all traces of NEW solution.

The DNA was recovered from the glass beads by adding 20 $\mu$ l of dH<sub>2</sub>O and incubating at 55°C for 5min. The glass beads were pelleted by centrifugation (12000g for 30s in a microfuge) and the supernatant retained. This final step was repeated using another 20 $\mu$ l of dH<sub>2</sub>O and the DNA solutions combined. The exact composition of the NaI, Glassmilk and NEW solutions were not disclosed by the manufacturers (Bio 101).

#### 2.3.12.2 ELECTRO-ELUTION

The TAE agarose gel was placed on a 302nm UV transilluminator and a trough measuring 0.5cm wide (the length and depth was determined by the length of slot and thickness of gel respectively) was excised from the TAE agarose gel immediately in front (with respect to direction of migration) of the DNA fragments of interest. The distal edge of the trough (with respect to the DNA) was lined with dialysis tubing (pre-treated by boiling in 2% (w/v) sodium bicarbonate, 1mM EDTA) and electrophoresis continued until the desired DNA had eluted from the agarose and migrated across the trough to be held by the dialysis tubing. The polarity of the current was then reversed momentarily before being switched off and the DNA was removed from the surface of the dialysis tubing using a pipette. The recovered DNA was then extracted with phenol:chloroform and chloroform and precipitated with ethanol. To recover relatively undamaged DNA, agarose gel electrophoresis was performed in the absence of EtBr. The migration of the DNA was monitored by removing a small "reference" strip of agarose containing DNA standards and staining it in EtBr.

#### 2.3.13 ISOLATION OF DNA FROM POLYACRYLAMIDE GELS

The following procedure, which is a modification of the technique described by Maxam and Gilbert (1971), was used to isolate radioactive probes for mapping the 5'-termini of RNA transcripts.

**Reagents:** Elution buffer; 500mM ammonium acetate, 10mM magnesium acetate, 1mM EDTA (pH 8.0), 10% (w/v) SDS.

**Crush and Soak method:** The DNA fragment to be isolated was resolved on a denaturing 6% (w/v) polyacrylamide.urea gel and located by autoradiography. A section of the gel containing the fragment of interest was removed using a scalpel blade and transferred to a microfuge tube. Using a disposable plastic pipette tip, the gel slice was crushed against the wall of the tube. The probe was eluted by adding 600µl of elution buffer and incubated at 37°C in an orbital shaker (contained within a lead pig) overnight. Fragments of acrylamide were removed using centrifugation to force the elutant through a syringe barrel packed with siliconised glass wool. The syringe was washed with 200µl of fresh elution buffer and the elutant and washings combined. The elutant was extracted twice with phenol.chloroform and twice with chloroform. The probe was then precipitated by adding 2 volumes of ethanol and harvested by centrifugation in a microfuge (12000g at 4°C for 30min). From this stage onwards RNAase free solutions and disposables were used. The pellet of single-stranded probe was rinsed once with 70% (v/v) ethanol and redissolved in 200µl DEPC-treated dH<sub>2</sub>O.

#### 2.3.14 TRANSFER OF DNA FROM *E. COLI* COLONIES TO AMERSHAM HYBOND-N<sup>TM</sup> MEMBRANES

**Reagents:** 20X SSC; 3M NaCl, 0.3M Tri-sodium citrate. Denaturing solution; 1.5M NaCl, 0.5M NaOH. Neutralising solution; 1.5M NaCl, 0.5M Tris.HCl (pH 7.2), 1mM EDTA.

**Procedure:** A Hybond-N<sup>TM</sup> membrane was cut to the correct size and placed onto the surface of a L-agar plate containing the appropriate antibiotic. The *E. coli* clones to be screened were streaked out on top of the membrane and the plate incubated at 37°C overnight. The membrane was then removed and placed, colony side up, on a pad of absorbent filter paper soaked in denaturing solution and left for 7min. Next the membrane was transferred to a pad of filter paper soaked in neutralising solution and left for 3min. This step was then repeated, with a fresh pad soaked in neutralising solution. The membrane was washed in 2X SSC, transferred to dry filter paper and air dried, colony side up. Finally, the samples were fixed to the membrane by baking in an oven at 80°C for 2 hours.

### 2.3.15 SOUTHERN ANALYSIS

Southern analysis was carried out by a method adapted from Southern (1975), and described in "Blotting and hybridization protocols for Hybond-N<sup>TM</sup> membranes" published by Amersham International plc.

#### 2.3.15.1 CAPILLARY TRANSFER

After electrophoresis and photography, the resolved DNA fragments were transferred under alkali conditions to Hybond-N<sup>TM</sup>.

**Reagents:** Denaturing solution; 1.5M NaCl, 0.5M NaOH. Alkali transfer buffer; 0.25M NaOH, 1.5M NaCl. 20X SSC; 3M NaCl 0.3M tri-sodium citrate.

**Procedure:** The gel was rinsed in distilled water, placed in enough denaturing solution to immerse it completely and left for 30min. The gel was removed, excess liquid removed by blotting and equilibrated for 10-15min in alkaline transfer buffer. The DNA was then transferred to the nylon membrane in alkali buffer, by capillary action (disposable nappies proved a particularly useful absorbent material for driving the transfer process). After blotting for a least 4 hours (but usually overnight), the membrane was washed briefly in 2X SSC to remove any adhering agarose.

#### 2.3.15.2 AQUEOUS PREHYBRIDIZATION AND HYBRIDIZATION CONDITIONS

The procedures which follow were suitable for DNA immobilised on filters from colony lifts and Southern transfer.

##### 2.3.15.2(a) HYBRIDIZATION USING RANDOM PRIMED PROBES.

**Reagents:** 100X Denhardt's solution; 10g ficoll, 10g polyvinylpyrrolidone, 10g bovine serum albumin. Made up to 500ml with dH<sub>2</sub>O. 20X SSC; 3M NaCl, 0.3M tri-sodium citrate.

**Prehybridization and hybridization conditions:** The prehybridization and hybridization solutions used in this work for the probing of homologous sequences contained 5X Denhardt's solution, 1X SSC, 0.5% (w/v) SDS (Sigma, Molecular Biology Grade) and 100µg.ml<sup>-1</sup>

sheared and denatured Salmon sperm DNA. The volume of the hybridization solutions depended on the size of the filters; 200 $\mu$ l were added for every cm<sup>2</sup> of membrane. The solutions were filtered through 0.4 $\mu$ m Nalgene membranes prior to the addition of Salmon sperm DNA and random primed probe (50ng, ca.  $1 \times 10^9$  cpm. $\mu$ g<sup>-1</sup>) (see section 2.3.9.1). The prehybridization and hybridization reactions were incubated at 65°C for 4 and 16 hours respectively.

**Washing of membranes after hybridization:** The nylon membrane was washed twice in 200ml of 1X SSC, 0.5% (w/v) SDS at 65°C for 20min each. Excess fluid was removed by blotting, the filter sealed in a plastic bag and subjected to autoradiography. The probe could be stripped from the filter (provided the filter was kept moist after hybridization) by washing in 0.4M NaOH at 45°C for 30min. The filter was then washed in 0.1X SSC, 0.2M Tris.HCl (pH 7.4) and 0.1% (w/v) SDS for 15min and stored in a sealed plastic bag until required.

#### 2.3.15(b) CONDITIONS FOR AQUEOUS HYBRIDIZATIONS WITH OLIGONUCLEOTIDES

This procedure was used for probing homologous DNA, which was immobilised on Hybond-N<sup>TM</sup> membranes, with radioactively labelled oligonucleotides of 20-30 bases.

**Prehybridization and hybridization conditions:** The surface area of the filter was measured and 200 $\mu$ l of prehybridization and hybridization used for every square centimetre. The prehybridization solution contained 6X SSC, 0.05% (w/v) SDS (Sigma, Molecular Biology Grade), 0.05% (w/v) sodium pyrophosphate and 200 $\mu$ g.ml<sup>-1</sup> Heparin (Sigma, Grade 1). The prehybridization solution was heated to 65°C before the required amount of heparin was dissolved. The solution was then added to the filter and incubated at 65°C for 4 hours. The hybridization solution was the same as the prehybridization solution, except that it contained 0.5% (w/v) SDS and 50ng of probe (ca.  $1 \times 10^9$  cpm. $\mu$ g<sup>-1</sup>). The hybridization solution was prewarmed to 65°C before the heparin and 50ng of oligonucleotide probe were added (see section 2.3.9.2). The filter was incubated with the hybridization solution at 65°C for 90min.

**Washing of the membrane:** The membrane was washed in 150ml of 5X SSC, 0.05% (w/v) SDS for 15min at room temperature and then twice in 150 ml of the same wash solution at 60°C.

### 2.3.16 PREPARATION OF SINGLE-STRANDED M13 DNA

The single-stranded M13 templates were prepared as described in the "M13 Cloning/Dideoxy sequencing Instruction Manual" published by Bethesda Research Laboratories.

**Minipreparations:** A single M13 plaque was used to infect 1.5ml of 2X YT broth containing 15µl of an overnight culture of *E. coli* TG1. This was grown at 37°C for 5-6 hours with vigorous shaking, then transferred to a microfuge tube and harvested by centrifugation at room temperature for 5min. The supernatant, containing the phage particles, was recovered and respun. The supernatant (1ml) that remained was mixed with 200µl of a solution of 20% (w/v) PEG (6000)/2.5M NaCl and left to stand at room temperature for 15min to precipitate the phage particles. These were recovered by centrifugation at room temperature in a microfuge for 15min. The supernatant was discarded, the pellet respun and all traces of supernatant removed. The pellet was then resuspended in 100µl dH<sub>2</sub>O and extracted twice with phenol:chloroform and twice with chloroform. The DNA was then ethanol precipitated from the aqueous phase and recovered by centrifugation in a microfuge. The ssDNA was then washed with 70% (v/v) ethanol, dried *in vacuo* and redissolved in 20µl dH<sub>2</sub>O.

**Maxipreparations:** The same overall procedure was followed as for the "minipreparations", except that all the volumes were scaled up 20 fold. The cells from a 30ml culture were spun out (14000g for 2min), the supernatant recovered and respun as before. The supernatant (20ml) that remained was then precipitated with 5ml of 20% (w/v) PEG/2.5M NaCl for 10 min at room temperature and the phage harvested by centrifugation (14000g at 20°C for 15min). The phage pellet was resuspended in 1ml of dH<sub>2</sub>O, then reprecipitated and processed as for the minipreparations with the volumes scaled up accordingly.

### 2.3.17 DNA SEQUENCING

Dideoxy sequencing (Sanger *et al.*, 1975) was carried out on single-stranded M13 templates using a TAQuenase™ kit supplied by United States Biochemical Corporation (USB).

**Annealing template and primer:** In a 1.5ml microfuge tube, 1µg of single-stranded template, 1µl of sequencing primer (0.5pmoles) and 2µl of reaction buffer were combined and the volume made up to 13µl with dH<sub>2</sub>O. After denaturing at 85°C for 2min, the primer was annealed to the template by incubating at 45°C for 10min. After annealing was complete, the tube was placed on ice and used within ca. 2 hours.

**Labelling reaction:** To the annealed template-primer, the following were added on ice:

	<u>Volume (µl)</u>
7-deaza-dGTP labelling mix	2
5µCi of [ $\alpha$ - <sup>32</sup> P]dATP (800Ci.mmol <sup>-1</sup> )	0.5
Cloned <i>Taq</i> DNA polymerase (2 units)	2

The above were mixed and incubated at 45°C for 5min. To read sequences close to the primer annealing site a 1:5 dilution of the labelling mixture was used.

**Termination reactions:** Immediately after the labelling reaction was complete, 4µl aliquots were dispensed into four microfuge tubes containing one of the four 7-deaza-dGTP termination mixtures; ddATP, ddCTP, ddGTP and ddTTP, respectively. The contents were mixed as quickly as possible and incubated at 70°C for 5min. After the reactions were complete, they were allowed to cool to room temperature, 4µl of Stop loading buffer was added and then they were stored on ice until the sequencing gel was ready to load. The samples were heated to 70°C for 5-10 min, cooled rapidly on ice and loaded immediately onto the gel. Approximately 2-3µl of the sequencing products were loaded per track. The compositions of the buffers and nucleotide mixes were not disclosed by the manufacturers (USB).



### 2.3.18 TRANSCRIPT MAPPING USING COMPLEMENTARY SINGLE-STRANDED DNA PROBES

Mapping of the 5' terminus of RNA transcripts was undertaken using single-stranded probes derived from M13 (Calzone *et al.*, 1987), as described by Maniatis *et al.* (1989).

**Reagents:** 1X Hybridization buffer; 40mM PIPES adjusted to pH 7.4 with NaOH, 400mM NaCl, 1mM EDTA, 80% (v/v) formamide (de-ionized with mixed-bed resin). Stored at -70°C. 10X S1 nuclease digestion buffer; 2.8M NaCl, 500mM NaCH<sub>3</sub>COO (pH 4.5), 45mM ZnSO<sub>4</sub>. Stored at -70°C. 10X Exonuclease VII digestion buffer; 500mM Tris.HCl (pH 7.8), 500mM KCl and 100mM EDTA (pH 8.0). Denatured Salmon sperm DNA; prepared by dissolving in dH<sub>2</sub>O, extracting any protein with phenol.chloroform, shearing by passage through a narrow gauge hypodermic needle and ethanol precipitating. The fragmented DNA was then redissolved in dH<sub>2</sub>O to a concentration of 10mg.ml<sup>-1</sup>, boiled for 10min and stored at -20°C in small aliquots. Prior to use, the DNA was partially denatured by heating to 100°C for 5min and chilling rapidly on ice. Carrier tRNA; prepared by dissolving commercially available yeast tRNA in dH<sub>2</sub>O, extracting any protein with phenol.chloroform, precipitating and redissolving in dH<sub>2</sub>O to a final concentration of 10mg.ml<sup>-1</sup>. S1 nuclease stop buffer; 500mM Tris.HCl (pH 8.0), 125mM EDTA.

**Probe synthesis:** 0.5 pmoles of template ssDNA was mixed with 3 pmoles of an oligonucleotide primer in a volume of 16  $\mu$ l, containing 2 $\mu$ l 10X BRL REact 2 buffer, and then annealed by heating to 85°C for 5min and cooling slowly to 37°C. The extension reaction was then carried out at 37°C for 30 min by adding:

	<u>Volume (<math>\mu</math>l)</u>
100mM dithiothreitol	2
50 $\mu$ Ci of [ $\alpha$ - <sup>32</sup> P]dATP (sp. act. 3000 Ci.mmmole <sup>-1</sup> )	5
40 $\mu$ M solution of dATP	1
20mM solution of dCTP, dGTP and dTTP	1
Klenow (1 U. $\mu$ l <sup>-1</sup> )	5

The above reaction was then "chased", by adding 1 $\mu$ l of a 20mM dATP solution and incubating for a further 20 min. After synthesis, the reaction was heated to 68°C for 10min to inactivated the Klenow. The probe was then digested at a restriction site, within or beyond the insert, by the addition of 20 $\mu$ l of restriction digest solution containing the appropriate restriction endonuclease (20U) and buffer. The reaction was incubated for 2 hours at the optimum temperature for the enzyme. The nucleic acid was then precipitated with ethanol and recovered by centrifugation in a microfuge. The DNA pellet was washed with 70% (v/v) ethanol and dried *in vacuo*. To purify the single-stranded probe, the pellet was resuspended in 20 $\mu$ l 1X sequencing loading buffer, denatured by heating to 85°C for 5min and resolved by electrophoresis on a denaturing 6% (w/v) polyacrylamide gel. The position of the probe was determined by autoradiography and eluted from the gel using the crush and soak protocol (see Section 2.3.13).

**Incorporation of radio-labelled nucleotides:** When the probe was used in quantitative S1 nuclease protection experiments, the percentage of radio-labelled nucleotides that were incorporated was calculated. 0.5 $\mu$ l samples removed from the extension mixture before the addition of Klenow and after the synthesis of the probe but before the "cold" chase were spotted onto 1 cm<sup>2</sup> square of Whatman GF/C filters in microfuge

tubes (1.5ml). The filters were stored at room temperature until all the fluid had evaporated. Using blunt-ended forceps, the filter containing the sample removed immediately before the "cold" chase was swirled in a beaker containing 100-150ml of ice-cold 5% (w/v) trichloroacetic acid and 20mM sodium pyrophosphate for 2min. This washing step was repeated three times using fresh volumes of the acid solution. The filter was then washed in 70% (w/v) ethanol for 30s and allowed to dry at room temperature. The two filters were then placed in separate scintillation vials and the amount of radioactivity measured by Cerenkov counting. The proportion of radio-labelled precursors and specific activity of the probe were calculated using the following equations:

$$\frac{\text{cpm in washed filter}}{\text{cpm in unwashed filter}} = \text{proportion incorporated}$$

$$\text{proportion incorporated} \times \text{total weight} = \text{total amount of product}$$

$$\frac{\text{cpm incorporated}}{\text{total amount of product}} = \text{specific activity}$$

This procedure yielded routinely 0.1 to 0.4 pmoles of single-stranded probe with a specific activity of  $10^6$ - $10^7$  Cerenkov cps.pmole<sup>-1</sup> DNA.

**Hybridization:** The single-stranded probe (ca. 0.1-0.2pmoles) and the *Streptomyces* RNA (5-20µg) were co-precipitated, washed with 70% (v/v) ethanol and dried by leaving the tube open on the bench, until the last traces of ethanol had evaporated. The probe/RNA pellet was resuspended in 20µl of 1X Hybridization buffer. The solution was pipetted up and down many (20-30) times and heated to 60°C for 30min to ensure that the pellet was dissolved completely. The hybridization solution was then heated to 85°C for 10min, to denature the nucleic acids and incubated at 55-65°C overnight. After overnight incubation, the hybridization mixture was either treated with S1 nuclease or Exonuclease VII.

**S1 Nuclease digestion:** 280µl of ice-cold 1X S1 digestion buffer, containing 10µg of partially denatured Salmon sperm DNA, was added to the hybridization reaction and mixed by inverting the microfuge tube several times. As quickly as possible, 250 units of S1 nuclease (1-2µl) were added, the reaction mixture pooled by centrifugation and incubated at 37°C for 30min. The reaction was terminated by the addition of 75µl stop buffer and 10µg carrier tRNA. The reaction mixture was then extracted with phenol:chloroform and chloroform, ethanol precipitated, washed with 70% (v/v) ethanol, dried *in vacuo* and resuspended in 7µl 1X sequencing loading buffer. The products were run against sequencing ladders on denaturing polyacrylamide gels.

**Exonuclease VII digestion:** 280µl of ice cold 1X exonuclease buffer, containing 10 units of exonuclease VII, were added to the hybridization solution, mixed by inverting the tube several times, pooled by brief centrifugation and incubated at 37°C for 45min. The reaction mixture was extracted with phenol:chloroform and chloroform and processed as for the S1 digestion products above.

### 2.3.19 PRIMER EXTENSION MAPPING OF TRANSCRIPT 5' ENDS

The protocol described below was communicated by Lewis Wray (Boston University, USA).

**Reagents:** 5X Annealing buffer; 1.0M KCl, 0.1M Tris.HCl (pH 8.3). 5X Elongation buffer; 0.9M Tris.HCl (pH8.3), 0.1M MgCl<sub>2</sub>.

**Procedure:** 5'-labelled oligonucleotide (ca. 0.5-1ng) and *Streptomyces* RNA (10-40µg) were co-precipitated, washed with 70% (v/v) ethanol and resuspended in 20µl 1X Annealing buffer. The sample was incubated at 80°C for 5min to denature and then incubated at 42°C for 3 hours. After the RNA-oligonucleotide annealing reaction was completed the following were added:

	<u>Volume (µl)</u>
5X Elongation buffer	2
Actinomycin D (1 mg.ml <sup>-1</sup> )	1
2mM dATP, dCTP, dGTP, dTTP solution	1
RNAguard Ribonuclease inhibitor (Pharmacia, 37U.µl <sup>-1</sup> )	1
Avian Myeloblastosis Virus (AMV) reverse transcriptase (Pharmacia)	1 (25 units)
100mM dithiothreitol	2
dH <sub>2</sub> O	2

The elongation reaction was incubated at 42°C for 30min and then terminated by the addition of 10µl Sequencing Stop loading buffer (USB, TAQuenase™ kit). Immediately prior to loading, the extension products were heated to 85°C. 3-5µl of the sample was then loaded onto the sequencing gel.

### 2.3.20 AUTORADIOGRAPHY

Autoradiography was performed in metal cassettes (medical chest X-ray type) using Kodak X-OMATS film. Enhancement in the intensity of the autoradiographic images was obtained when the film was exposed at low temperature (-70°C) in close contact with a du Pont Cronex Lightening Plus intensifying screen. The X-ray films were developed using a Kodak X-OMAT automatic processor, Model ME-I.

### 2.3.21 DENSITROMETRY SCANNING

The intensity of autoradiographic images were determined using a Hoefer Scientific Instruments GS-300 Scanning Densitometer, integrated with an IBM GS-360 Data System.

---

## **CHAPTER 3**

### **ANALYSIS OF PROMOTER ACTIVITY UPSTREAM OF *OTRA***

### 3.1 Introduction

The biosynthetic genes for oxytetracycline production are located in a single cluster, flanked by two resistance genes (Binnie *et al.*, 1989). Both the *otrA* and *otrB* resistance genes have been cloned from *S. rimosus* M15883, a commercial production strain developed by Pfizer (Rhodes *et al.*, 1984). The product of the *otrA* resistance gene protects the ribosomes from translational arrest by oxytetracycline, while the *otrB* gene product is involved in efflux of the antibiotic from the mycelium (Ohnuki *et al.*, 1985; Doyle *et al.*, 1986). During vegetative growth, tetracycline resistance in *S. rimosus* can be induced by pre-exposure to sub-lethal concentrations of tetracycline (Ohnuki *et al.*, 1985; Lee, unpublished result).

The *otrA* resistance gene has been sequenced and the pattern of transcription analysed partially (Figure 3.1; from the results of Doyle, 1987). Two transcription start points were proposed for *otrA*, 129nt and 339-344nt upstream of the translation initiation codon, corresponding to the promoters *otrAp1* and *otrAp2* respectively. Transcription has been shown to terminate immediately downstream of the *otrA* resistance gene at a site which is predicted to form a typical rho-independent terminator structure in the messenger RNA (Doyle, 1987). The recognition sequence of *otrAp1* resembles the consensus sequence for the major class of eubacterial promoters. This promoter is thought to be transcribed from the *S. rimosus* equivalent of the *S. coelicolor* RNA polymerase holoenzyme, E $\sigma^{35}$ , which is capable of transcription from the *veg* promoter of *B. subtilis* and the *dagAp4* from *S. coelicolor* (General Introduction, 1.9) (Westpheling *et al.*, 1985; Buttner *et al.*, 1988). No DNA sequence was obtained for *otrAp2*, as it was external to the cloned DNA conferring tetracycline resistance (General Introduction, 1.14).

Owing to the commercial importance of *S. rimosus* M15883, *in vivo* manipulations can only be carried out within the confines of the Pfizer Laboratories. Consequently, previous analyses of *otrA* transcription used RNA isolated from the ancestral production strain *S. rimosus* M4018, grown on Tryptone Soya Broth (TSB), which does not support antibiotic production (Doyle, 1987). As no sequence resembling a rho-

independent terminator could be detected in the intervening region between *otrA* and *otcZ* (a biosynthetic gene located immediately upstream; Figure 3.1), it was proposed that these genes may be co-transcribed during antibiotic production, as part of a polycistronic message.

Without exception, antibiotic production genes have been found in clusters. Indeed, where all the biosynthetic genes have been identified or where heterologous expression of antibiotic production has been attained, all of the biosynthetic genes and at least one resistance gene have been mapped to a single location (Malpartida and Hopwood, 1984; Stanzak *et al.*, 1986; Chen *et al.*, 1986; Motamedi and Hutchinson, 1987; Binnie *et al.*, 1989). The close proximity of resistance determinants to biosynthetic genes and the requirement for their concomitant expression during physiological differentiation has prompted speculation that the expression of these genes may be regulated coordinately (Hopwood, 1988).

### 3.1.1. Objectives

The transcriptional analysis of *otrA* promises to reveal interesting facets about the coordination of gene expression during the switch to secondary metabolism and the induction of antibiotic resistance during vegetative growth.

The immediate aims of the experiments discussed in this chapter were as follows.

(i) To identify the promoters responsible for the transcription of the *otrA* resistance gene and to compare their recognition sequences with the consensus sequences for known classes of promoters.

(ii) To establish culture conditions for *S. rimosus* M4018 and M15883 that yielded reproducible growth and oxytetracycline production profiles.

(iii) To isolate RNA from *S. rimosus* cultures at different stages in the fermentation of oxytetracycline and to determine the temporal changes in the transcription of *otrA*.



(iv) To identify suitable RNA samples/ time points for intensive transcriptional analyses of the OTC cluster.

### 3.1.2 Methodology

Two strategies were employed to investigate the transcription of *otrA*; (i) the cloning of fragments from the region immediately upstream of *otrA* into a promoter-probe system and (ii) high-resolution S1 nuclease protection experiments.

At the onset of this investigation (1987) only two plasmid systems were available readily for measuring *in vivo* promoter activity in *Streptomyces*; pARC1, a low-copy-number vector (Horinouchi and Beppu, 1985) and pIJ486/7, two high-copy-number vectors that differ only in the orientation of their multiple cloning sites (Ward *et al.*, 1986).

It has been known for some time that the molar ratio of transcriptional regulators to their recognition sites on DNA is of major importance in regulating gene expression (Ptashne, 1986). As the pARC1 system is based on a low-copy-number plasmid, it should be useful for studying promoters that are regulated by the binding of ancillary proteins. The pARC1 plasmid contains a promoterless locus specifying the synthesis of a brown pigment. Recombinants containing a promoter, which is actively directing the transcription of the pARC1 reporter locus, have a phenotype which can be detected easily without selection. This system can be used, therefore, to analyse promoters with distinct temporal activities. Unfortunately, pARC1 has only a single *Bam*HI site, upstream of the reporter locus, which is suitable for inserting fragments to be tested.

The pIJ486/7 system includes several useful features; a multiple cloning site located immediately upstream of the *aphII* reporter gene, which encodes kanamycin resistance, an *E. coli* bacteriophage fd terminator upstream of the cloning sites to prevent transcriptional read-through from vector promoters and a translational stop codon, which is located 3' to the cloning site but in frame and 5' to the reporter gene,

to avoid translational fusions, which could alter the specific activity of the *aphII* gene product. Although the high copy-number of the pIJ486/7 plasmids permits the detection of weak promoters, any regulation dependent on the ratio of target DNA to ancillary protein may be disrupted.

As *S. rimosus* M4018 and M15883 produce a brown pigmented by-product during the production of oxytetracycline and are inherently resistant to high levels of kanamycin (ca. 500 $\mu$ g.ml<sup>-1</sup>, Hunter pers. comm.), neither the pARC1 nor the pIJ486/7 system can be used to measure promoter activity in these production strains. Previously, the pIJ486/7 plasmids were used to analyse the *in vivo* activity of *otrA*<sub>p1</sub>, due to the ease with which DNA fragments could be inserted in a specific orientation (Doyle, 1987). The assays were carried out using the heterologous host *S. lividans*, which is highly sensitive to kanamycin and can be transformed with high efficiency. In order that measurements of promoter activities could be compared directly, this approach was also adopted to analyse further the transcription of *otrA*.

The S1 nuclease protection assay (Berke and Sharp, 1977) is the most widely used technique for determining the precise locations of the 5' and 3' termini of transcripts. Several variations on the original technique have been published (for an overview see Calzone *et al.*, 1987). Single-stranded probes derived from recombinants of bacteriophage M13 were used to map the 5' termini of the *otrA* transcripts (as described by Maniatis *et al.*, 1989). This approach has several advantages; (i) the purification of single-stranded probe minimises the possibility of the complementary DNA strand (i.e the template) competing with RNA for hybridization to the probe, (ii) the high-specific activity of the probes, which can exceed 10<sup>9</sup> Cerenkov cpm. $\mu$ g<sup>-1</sup> DNA, facilitates the detection of rare transcripts and (iii) the templates, which are used to produce the probe, can also be used to generate dideoxy-mediated sequencing ladders for the accurate sizing of the nuclease-resistant fragment(s).

## 3.2 Results

### 3.2.1 Construction of Recombinants for *in vivo* Promoter Analysis.

DNA fragments from the region upstream of *otrA* (Figure 3.1) were subcloned into the pIJ486 promoter probe vector (Ward *et al.*, 1986). The copy numbers of these constructs were not significantly different, as judged by the relative yields from numerous plasmid preparations. Initially, the 305bp *Bcl*I5-*Kpn*I7 (isolated as a *Kpn*I fragment using a site in the polylinker), the 722bp *Kpn*I7-*Pst*I8 and the 805bp *Bam*HI6-*Pst*I8 fragments from pPFZ513 (Binnie, Pfizer) were subcloned into pUC18 to produce pKME805, pKME807 and pKME808 respectively. The *Bcl*I5-*Kpn*I7 insert was orientated such that the *Bcl*I site was nearest the *Eco*RI site in the polylinker. This intermediate cloning step was undertaken as *E. coli* recombinants could be analysed rapidly and the subsequent subcloning step into the streptomycete vector was much simpler, that is the same preparation of *Hind*III/*Eco*RI-digested pIJ486 could be used for all three cloning experiments.

*Hind*III/*Eco*RI fragments from the pUC18-based constructs were then ligated into pIJ486, which had been treated with calf intestinal phosphatase, to derive pKMS605 from pKME805, pKMS607 from pKME807 and pKMS608 from pKME808 respectively. The products of each ligation were then introduced into *S. lividans* TK54 by genetic transformation. For each transformation, plasmid DNA was isolated from at least three different thiostrepton-resistant transformants and digested with restriction endonucleases. Although gross rearrangements were never detected, occasionally, plasmids were found which did not contain an insert. *Hind*III/*Eco*RI and *Pst*I digests of plasmid DNA from the isolates that were assayed subsequently for *in vivo* promoter activity are shown in Figure 3.2. All of these constructs produced the expected band patterns when the products of endonucleolytic cleavage were separated by electrophoresis on agarose gels. The inserts were orientated such that any promoter activity capable of directing transcription of *otrA* would promote expression of the *aph*II reporter gene.

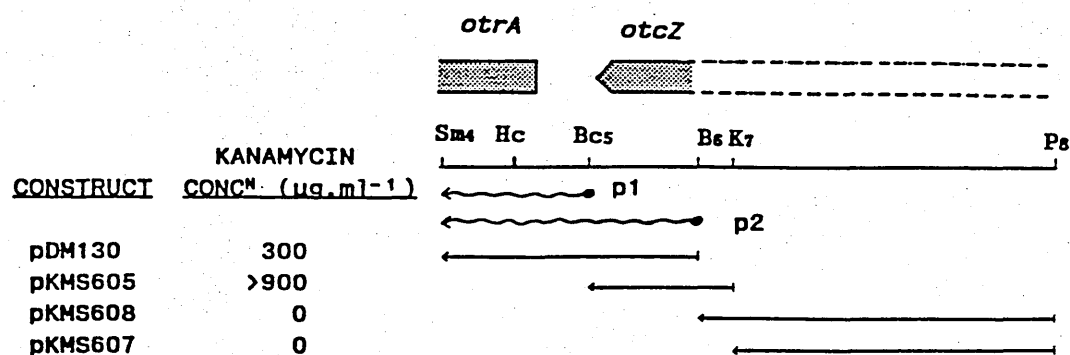
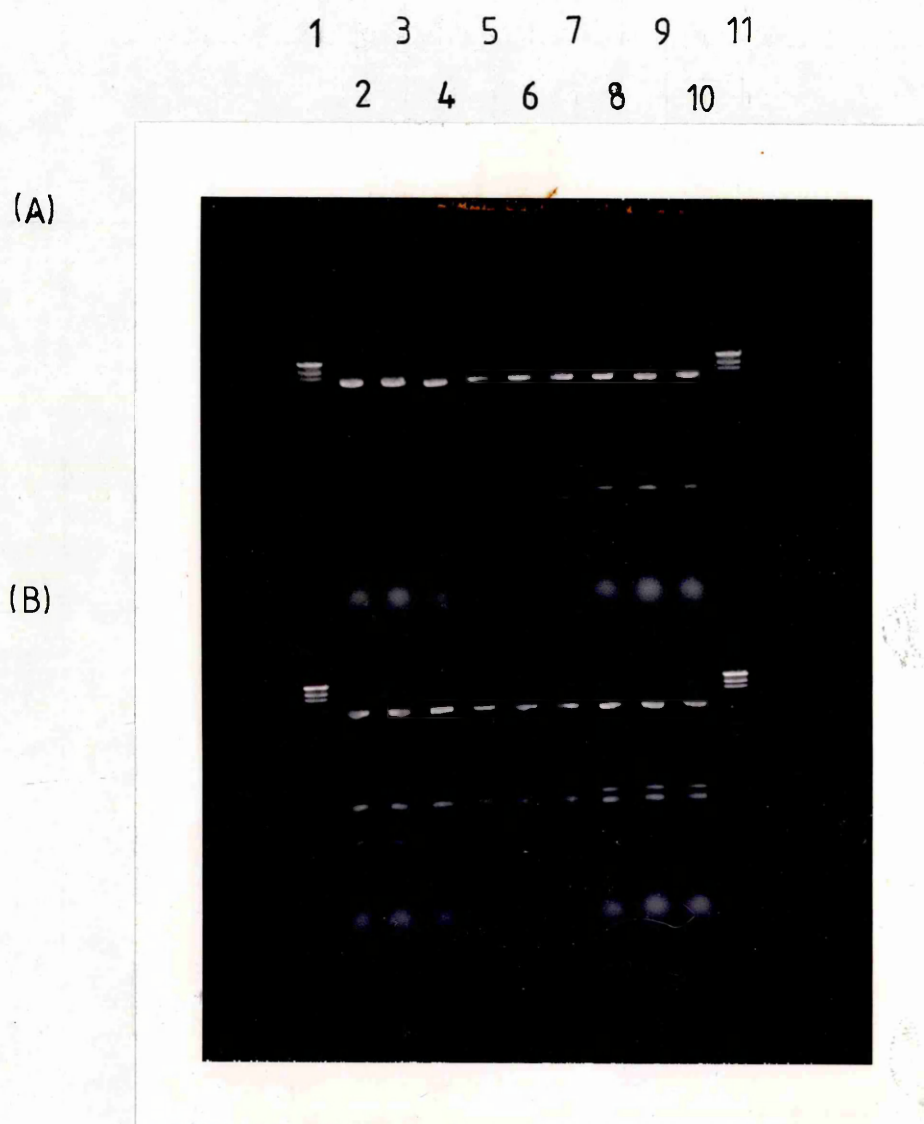


Figure 3.1 Promoter-probe constructs used to investigate promoter activity capable of transcribing the *otrA* resistane gene. Arrows indicate the orientation of the insert with respect to the promoterless *aphII* gene in pIJ486. Shaded boxes (■) denote the *otrZ* and *otrA* protein coding regions. Transcripts detected by Doyle (1987) are represented by bold wavy arrows, the closed circles indicating the proposed start site. The level of kanamycin conferred by the constructs is indicated in the column at the left of the figure. Abbreviations: B, *Bam*HI; Bc, *Bc*II; Hc, *Hinc*III; K, *Kpn*I; P, *Pst*I; Sm, *Sma*I.



**Figure 3.2.** Restriction digests to confirm the status of the inserts in independent isolates of the promoter-probe constructs.

Gel A; *Hind*III/*Eco*RI digests of the pIJ486 constructs

Gel B; *Pst*I digests of the pIJ486 constructs

Lanes 1 and 11; lambda *Hind*III DNA markers

Lanes 2-4; pKMS605 from isolates 1, 2 and 3

Lanes 5-7; pKMS607 from isolates 1, 2 and 3

Lanes 8-10; pKMS608 from isolates 1, 2 and 3

### 3.2.2 Assay of Promoter Activity from pIJ486 Constructs

Serial dilutions of spore suspensions of *S. lividans* TK54 containing each of the above constructs were spread out on Emersons agar plates to determine the spore titre. Similarly, spore suspensions of transformants containing pIJ486 and pDM130 (a pIJ486 construct containing the *Sma*I<sub>4</sub>-*Bam*HI<sub>6</sub> fragment; Doyle, 1987) were also prepared as negative and positive controls respectively.

To assay promoter activity, approximately  $10^2$  spores of each of the above transformants were streaked out on Emersons agar plates containing thiostrepton ( $25\mu\text{gml}^{-1}$ ) and kanamycin at concentrations ranging from  $100\mu\text{gml}^{-1}$  to  $900\mu\text{gml}^{-1}$ . Growth on the agar plates was assessed after incubation for 7 days at  $30^\circ\text{C}$  (Figure 3.1). The only recombinants to exhibit promoter activity were *S. lividans* TK54 containing pKMS605 and the positive control pDM130. Using RNA isolated from *S. lividans* TK64 containing pDM130, the 5' endpoints of transcripts originating from *otrAp1* had been mapped at 220 and 221nt to the left of the *Bam*HI<sub>6</sub> site, within the *Bcl*I<sub>5</sub> recognition sequence (Doyle, 1987). Similar levels of promoter activity were also detected using pDM110, a pIJ486-based construct containing the 221 bp *Bcl*I<sub>5</sub>-*Bam*HI<sub>6</sub> fragment (Doyle, 1987). In addition to confirming the presence of transcripts originating from *otrAp1*, S1 nuclease mapping experiments using RNA from *S. rimosus* M4018 also detected a second signal, which was proposed to correspond to a transcript arising from a second promoter, designated *otrAp2*. The transcriptional start site for this promoter was mapped 5 to 10nt to the left of the *Bam*HI<sub>6</sub> site within the predicted protein coding region (PPCR) of *otcZ* (Doyle, 1987).

At this stage in the present analysis, the simplest explanation for the increased kanamycin resistance in recombinants containing pKMS605 ( $>900\mu\text{gml}^{-1}$ ) compared with those containing pDM130 (ca.  $300\mu\text{gml}^{-1}$ ) was that both *otrAp1* and *otrAp2* were directing transcription of the *aphII* reporter gene in the pKMS605 construct. The recognition sequence of *otrAp2* was interpreted as overlapping with the *Bam*HI<sub>6</sub> site, as the pKMS608 construct, which contains the *Bam*HI<sub>6</sub>-*Pst*I<sub>8</sub> fragment, did not confer kanamycin resistance to *S. lividans* TK54, even to concentrations

as low as 20 $\mu$ gml<sup>-1</sup>. The finding that the pKMS607 construct did not contain promoter activity was not unexpected as it contains neither *otrAp1* nor *otrAp2*.

A possible shortcoming of using the outgrowth of spore suspensions to assay for *in vivo* promoter activity in the pIJ486/7 system is that constructs containing promoters, which are not transcribed during vegetative growth, may not confer kanamycin resistance to germinating spores and consequently may not be detected. In an attempt to circumvent this possibility, kanamycin-resistance assays were also carried out using mycelial inoculums from liquid cultures of *S. lividans* TK54 containing pKMS607 and pKMS608 at various stages in growth, including stationary phase. Promoter activity could still not be detected and it was concluded that in the absence of any OTC-specific activator, the transcription of *otrA* was mediated solely by *otrAp1* and *otrAp2*.

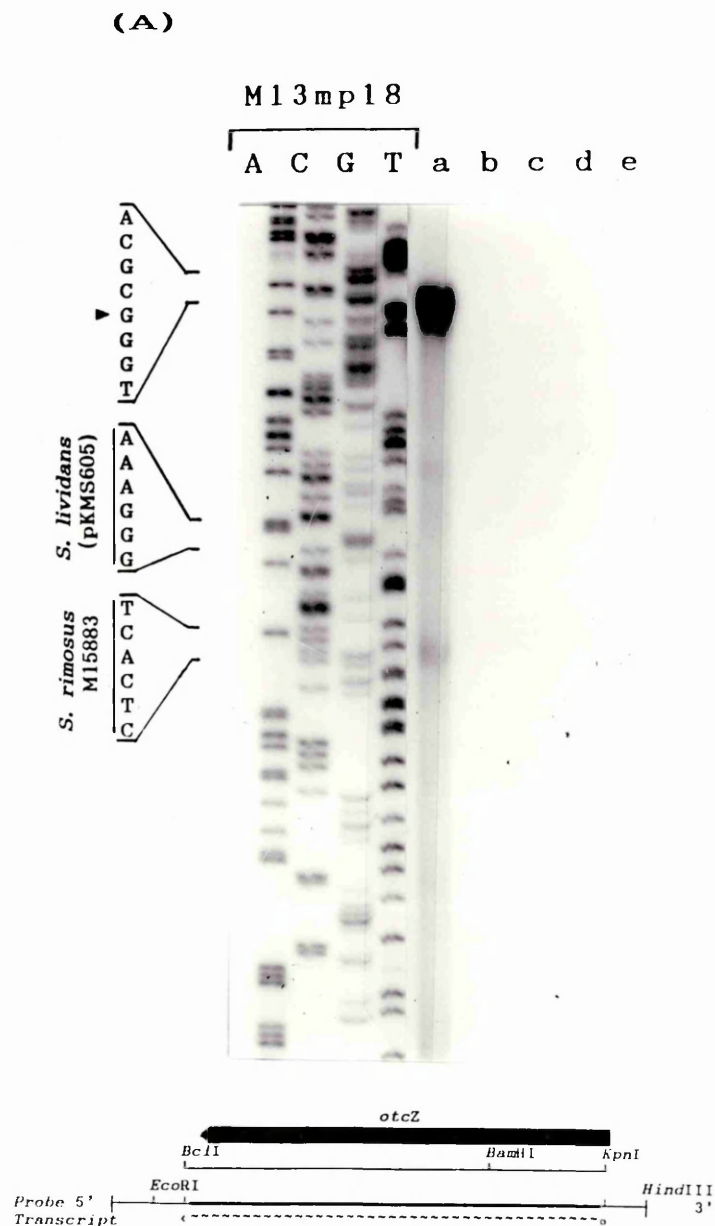
### 3.2.3 S1 Nuclease Mapping of the Putative *otrAp2* Transcripts

High resolution S1 protection experiments were undertaken to establish the precise 5' endpoint of the putative transcripts originating from *otrAp2*. Single-stranded probe, complementary to the anti-sense strand of *otrA*, was prepared from mKM905 (containing the *BcII*-*KpnI* fragment with the *BcII* site nearest the primer-annealing site, Chapter 4) by extension from the universal (-20) primer and digestion with *HindIII* (Materials and Methods; 2.3.18). This continuously-labelled probe was used in high resolution S1 mapping experiments with total RNA from three independent isolates of *S. lividans* TK54 containing pKMS605, *S. rimosus* M15883 (a Pfizer production strain) and *S. rimosus* M15883S (a spontaneous mutant with the entire OTC cluster deleted) as a control. Transcripts originating from 5-10nt to the left of the *BamHI* site were predicted to result in protected fragments of 212-217nt using RNA from *S. rimosus* M15883 and 233-238nt using RNA from *S. lividans* TK54 (pKMS605). The difference in the size of the predicted protection is due to the putative transcripts from pKMS605 having additional complementarity to polylinker sequences derived from M13mp19 in the probe. These transcripts should have been resolved easily in

denaturing polyacrylamide gels. However, no RNA-protected fragments were detected, which could correspond to the protection of probe by transcripts originating from around the *Bam*HI site (Figure 3.3, Panel A). Furthermore, repeated high-resolution S1 mapping experiments using numerous batches of RNA from *S. lividans* TK54 (pKMS605) failed to detect protection of any part of the probe (result not shown). As transcripts originating from *otrAp1* in the pKMS605 construct only have complementarity to 22-23nt of the polylinker sequence in the probe, it is possible that a RNA-protected fragment corresponding to *otrAp1* was not detected due to the poor stability of the probe-RNA complex under the conditions used for hybridization (Materials and Methods 2.3.18) and/or the low specific activity (cpm.pmol<sup>-1</sup>) of the short RNA-protected fragment.

Using RNA isolated from *S. rimosus* M15883 and M4018 (a former OTC production strain, Pfizer), a protected fragment of approximately 305nt was always detected (Figure 3.3, Panel B). The absence of a RNA-protected fragment using RNA from *S. rimosus* M15883S indicated that this protection was specific to the OTC cluster. As the probe contains nucleotides at both the 5' and 3' end that are derived from M13mp19, it is possible to distinguish between the fragment produced by reannealing of the probe to contaminating DNA template and probe protected by transcripts originating outwith but extending through the region cloned in M13mp19. The size of the protected band corresponded exactly to that predicted for transcripts extending through the entire 305bp *Bc*II-*Kpn*I region. In light of these S1 nuclease mapping experiments, the protection interpreted by Doyle (1987) as corresponding to transcripts originating from *otrAp2* is now considered to be full-length probe (FLP) which has had a few nucleotides removed from the 3' end by over digestion with S1 nuclease. Consideration of the combined results from the *in vivo* analysis of promoter activity and the S1 nuclease protection experiments, suggests that; (i) *otrAp1* is the only promoter capable of directing transcription of *otrA* within the *Bc*II-*Pst*I region and (ii) a promoter located at least 1.4kb upstream (beyond the *Pst*I site) also directs transcription of this resistance gene.





**Figure 3.3 (A).** High resolution S1 mapping of transcripts within the *Bcl*II-*Kpn*I region. Single-stranded probe ( $1 \times 10^6$  Cerenkov cpm.pmoles<sup>-1</sup>) was generated from mKM905 by extension from the universal primer and digestion with *Hind*III (see Materials and Methods, 2.3.18). 0.2 pmoles of probe were hybridized at 55°C with 10 µg of RNA from the following: (a) *S. rimosus* M18883, producing OTC (ca. 2 mg.ml<sup>-1</sup>) in trusoya medium 1 (TS1); (b) *S. rimosus* M15883S (a spontaneous mutant of M15883 deleted for the entire OTC cluster) grown on TSB; (c); (d) and (e) different isolates of *S. lividans* TK54 (pKMS605) grown on YEME containing thiostrepton (25 µg.ml<sup>-1</sup>). All samples were digested with S1 nuclease. An arrow indicates the position in the M13mp18 sequencing ladder coincident with the protected probe band in lane (a). Vertical lines (|) indicate the position at which probe protected by transcripts originating from *otrA*p2 were expected to migrate. Shown at the bottom of the panel is an interpretation of the S1 nuclease protection. The wavy arrow indicates the extent of probe protection by transcripts originating far upstream of *otrA*.



### 3.2.4 Transcription of *otrA* during OTC Production in Batch Culture

The temporal transcription of the *otrA* resistance gene during the production of oxytetracycline was investigated using RNA isolated from two different batch fermentations; (i) *S. rimosus* M4018 grown on Liquid Complete Medium (Figure 3.4) and (ii) *S. rimosus* M15883 grown on TS1 (a production medium developed by Pfizer; Figure 3.6). For clarity, the S1 nuclease protection analyses of RNA isolated from the different fermentations will be presented separately.

In each case, a single-stranded probe including 387nt of sequence from a *HincII* site between the *SmaI*<sub>4</sub> and the *BclI*<sub>5</sub> site to the *BamHI*<sub>6</sub> site (see Figure 3.1 for the location of restriction sites) was synthesized from mKM804 (Chapter 4) by extension from the universal (-20) primer and digestion with *EcoRI* (Materials and Methods, 2.3.18). Transcripts originating from *otrA*<sub>p1</sub> and far upstream protected DNA fragments of 166-167nt and 387nt respectively, which were resolved easily in denaturing polyacrylamide gels (Figures 3.5 and 3.7).

To calculate the relative abundance of transcripts originating from *otrA*<sub>p1</sub> and far upstream, it was assumed that radio-labelled nucleotides were evenly distributed along the entire length of the probe. Under standard conditions, the *HincII*-*BamHI* probes were synthesized from 0.5pmoles of mKM804 template in the presence of 16 pmoles of [ $\alpha$ -<sup>32</sup>P] dATP and 40 pmoles of "cold" dATP. The efficiency with which radio-label was incorporated was always greater than 65%, that is at least 36 pmoles of dATP were incorporated into the probes. As 0.5pmoles of template was present in the labelling reactions, the average probe would have incorporated at least 72 dATP nucleotides. As the average G+C content of streptomycete DNA is 73% mol (Enquist and Bradley, 1971), the primers should have been extended, on average, by 530 nucleotides. Therefore, the majority of the reaction products should have contained the double-stranded *EcoRI* site used to process the probe.

The transcription of *otrA* at the onset of oxytetracycline (OTC) production was examined using *S. rimosus* M4018 grown on Liquid Complete Medium, which produces typically 1-2mg.ml<sup>-1</sup> of OTC after 10-

12 days (Figure 3.4). RNA isolated from *S. rimosus* M4018, before and after the first time point at which OTC was first detected in the medium, was analysed by high-resolution S1 nuclease mapping using the *HincII*-*Bam*HI probe (Figure 3.5). The intensity of the bands produced by autoradiography was determined by densitometry scanning (Materials and Methods, 2.3.21). After correction for differences in the length of the radio-labelled fragments, the band corresponding to 0.01 pmoles of untreated probe in lane (a) was shown to be 2 to 10 fold more abundant than the probe bands protected by the *otrA* transcripts in samples (c) to (e). As the RNA samples were hybridized with 0.1 pmoles of probe, the molar excess of probe to *otrA* transcripts was calculated to be between 20 to 100 fold, which is within the limits recommended by Calzone and others (1987). RNA isolated from *S. rimosus* M4018 grown on TSB, which was analysed previously (Figure 3.3, Panel B), was used as a marker to indicate RNA-protected fragments corresponding to transcripts originating from *otrA*p1 and "far upstream" (lane b).

During the period of rapid growth, namely 24 hrs after inoculation and before the onset of OTC production, transcripts originating from *otrA*p1 accounted for greater than 95% of the total *otrA* transcription (Figure 3.5, Panel B). During the transition to OTC production, the relative abundance of the *otrA* transcripts changed dramatically. At 32 hours, when OTC was still not detected in the culture supernatant, the contribution made by *otrA*p1 had decreased to 80%. During the early phase of OTC production, 56 hours after inoculation, transcription from the promoter(s) located far upstream (pFAR) accounted for 51% of the total transcription of *otrA*. This shift in the relative contribution of p1 and pFAR to the transcription of *otrA* was accompanied by overall decrease of approximately 80% in the abundance of *otrA* transcripts within the total RNA population (Figure 3.5, Panel B).

The transcriptional analysis described so far, including that of Doyle (1987), has utilised subclones of DNA isolated from *S. rimosus* M15883 and RNA isolated from *S. rimosus* M4018. The strain development programme, which derived *S. rimosus* M15883 from *S. rimosus* M4018

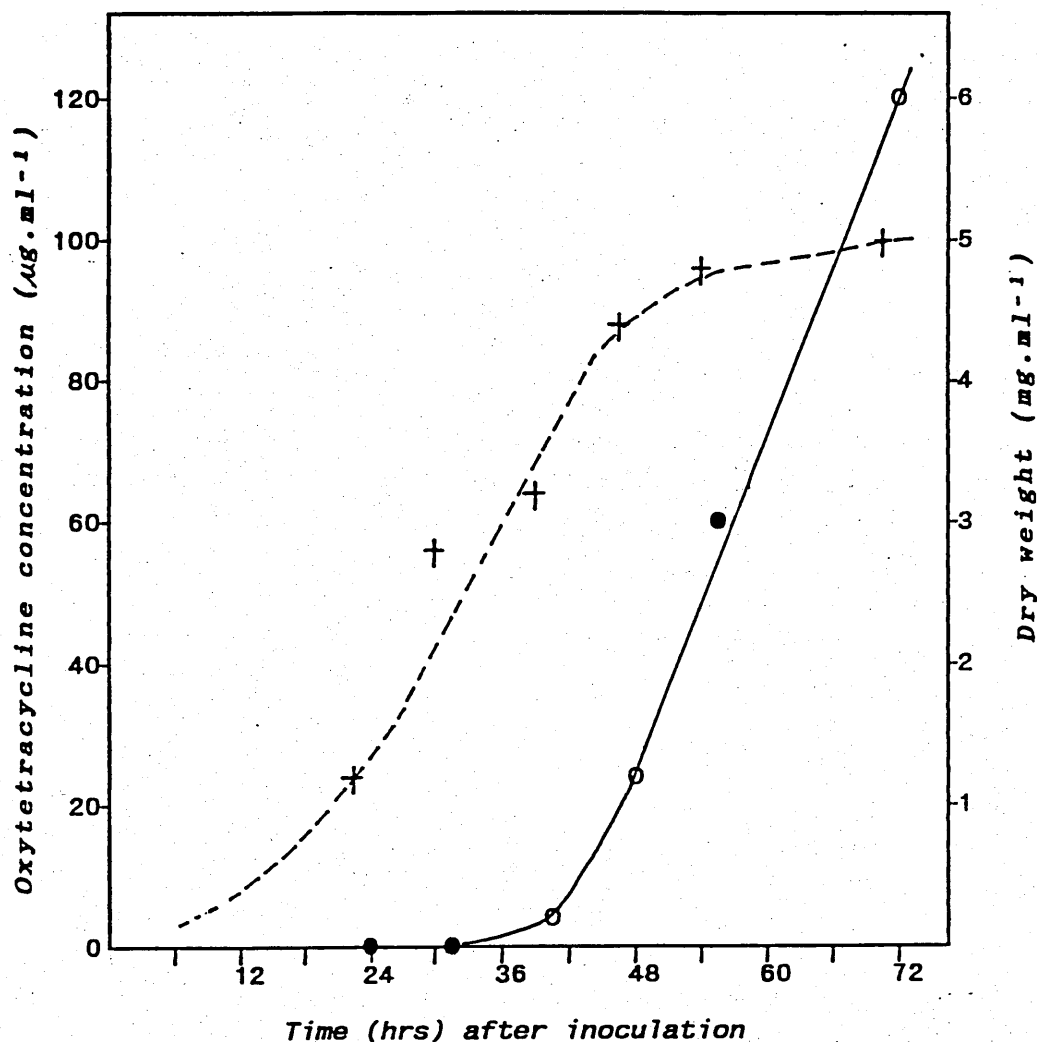
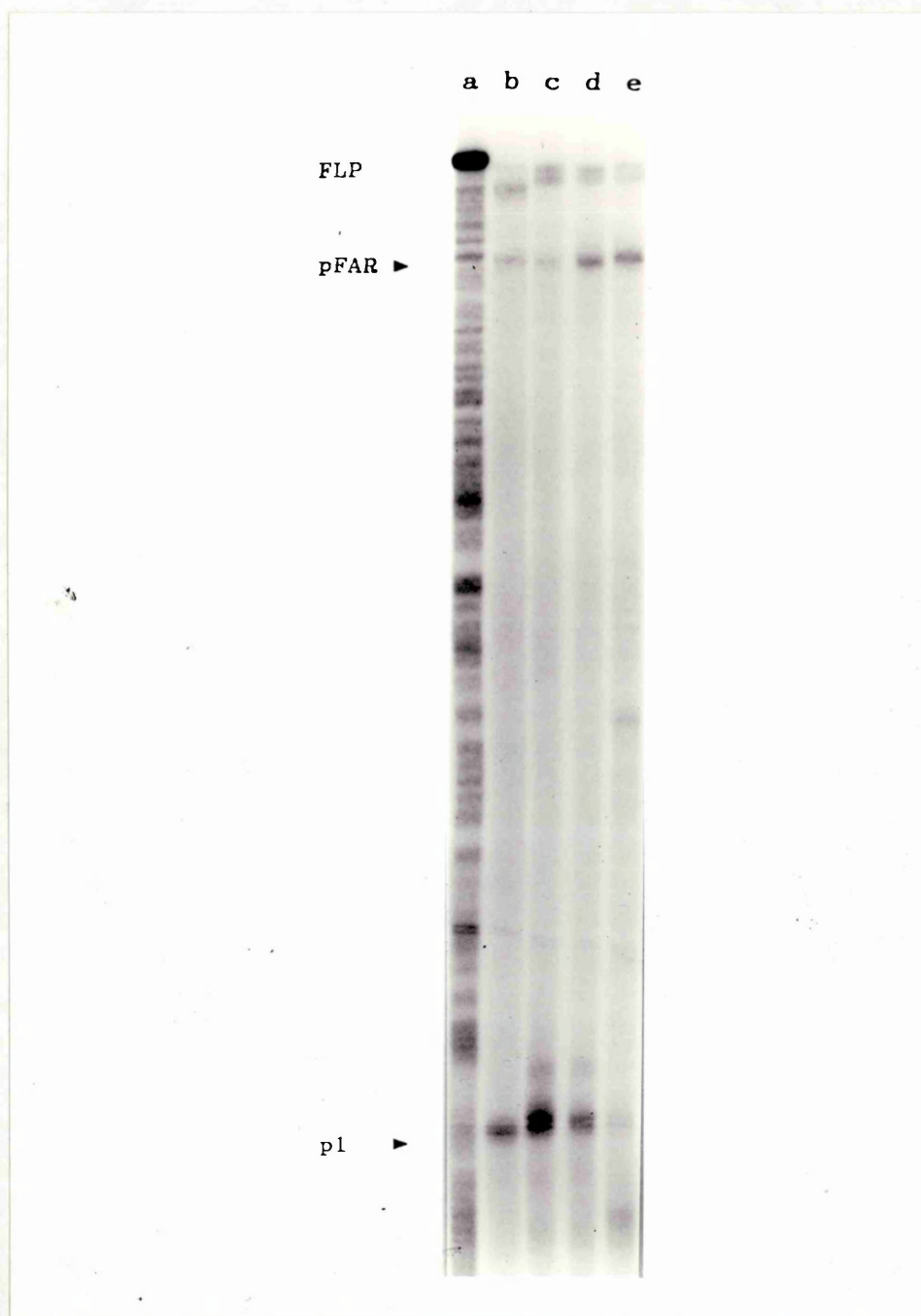


Figure 3.4. Oxytetracycline production by *S. rimosus* M4018 grown on Liquid Complete Medium. The graph is derived from 6 samples (O) taken from different cultures (50ml), inoculated at the same time with  $10^7$  spores and incubated in an orbital shaker at 30°C. The ● symbol indicates the concentration of OTC in cultures from which RNA was isolated and subsequently used in S1 nuclease protection experiments (Figure 3.5). The symbol + indicates the dry weight (mg.ml<sup>-1</sup>) of the mycelium at different stages in the fermentation.





**Figure 3.5. (A)** S1 nuclease mapping of *otrA* transcripts in total RNA isolated from *S. rimosus* M4018 grown on Liquid Complete Medium (LCM) (see Figure 3.4). Single-stranded probe (0.2 pmoles,  $2.1 \times 10^7$  Cerenkov cpm.pmol<sup>-1</sup>) synthesized from mKM804 by extension from the universal primer and digestion with *EcoRI* (see Materials and Methods, 2.3.18) was hybridized at 55°C with 10 µg of RNA from M4018 grown on LCM for 24, 32 and 56 hrs (lanes c, d and e). RNA isolated from M4018 grown on TSB was also included to provide markers for the RNA-protected fragments corresponding to transcription from *otrA*p1 and far upstream, which are indicated by p1 and pFAR respectively (lane b). The position of full-length probe (FLP) is also shown. 0.01 pmoles of untreated probe were also loaded on the gel (lane a).

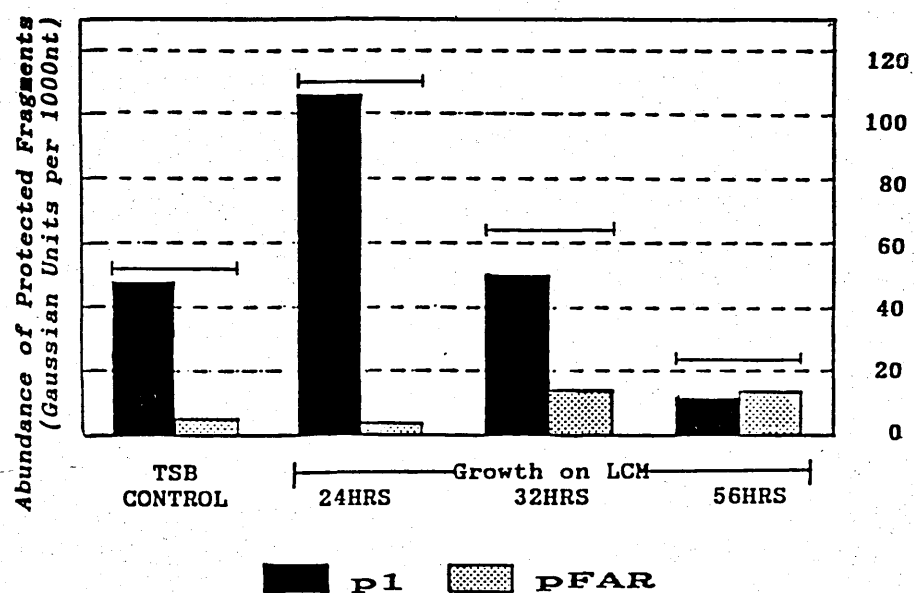


Figure 3.5 (B) Histogram showing the results of quantitative analysis of the *otrA* transcripts originating from p1 and far upstream in total RNA isolated from *S. rimosus* grown on LCM. Abundance estimates were obtained by densitrometry scanning of autoradiographs (Materials and Methods, 2.3.21) produced by different exposures to the denaturing polyacrylamide gel in which RNA-protected fragments had been resolved.

through several rounds of mutagenesis and selection for survivors that produced higher titres of OTC, has almost certainly lead to sequence differences within the OTC cluster of these two strains.

To obtain RNA that would be completely homologous with the DNA probes and could, thus, be used to analyse the transcription of the entire OTC cluster, a short period of time was spent at the Pfizer laboratories in Sandwich. Culture conditions were developed for *S. rimosus* M15883 that yielded reproducible oxytetracycline profiles, thus enabling RNA to be isolated at different stages during the production of OTC.

Aliquots (1ml) of a stationary phase culture of *S. rimosus* M15883 grown in Liquid Complete Medium (LCM) were used to inoculate TS1 production media (50 ml) in Erlenmeyer flasks (300 ml). The cultures were incubated on a rotary shaker with a 4cm stroke at 28°C. Samples of the supernatant were analysed frequently for OTC and RNA was isolated at different phases in the fermentation (Figure 3.6).

It was not possible to determine a growth profile for *S. rimosus* M15883, as the production medium (TS1) contained a colloidal suspension of soya flour and starch and an emulsion of vegetable oils, which interfered with the estimations of growth using optical density measurements and dry weight determinations. Furthermore, RNA could only be isolated after incubation for 18 hours, as the soya flour and starch, which were not utilised as substrates for growth during the early stages of the fermentation, formed an impervious "nougat-like" layer around the mycelium when they were harvested by either centrifugation or filtration.

High resolution S1 nuclease protection, as described in the previous section, was employed to analyse the RNA isolated from cultures of *S. rimosus* M15883 at different stages during the production phase (Figure 3.7; samples d to g). The most striking features of this analysis were (i) the absence of detectable amounts of transcripts from *otrA*<sub>p1</sub> and (ii) the transient peak in the abundance of *otrA* transcripts from



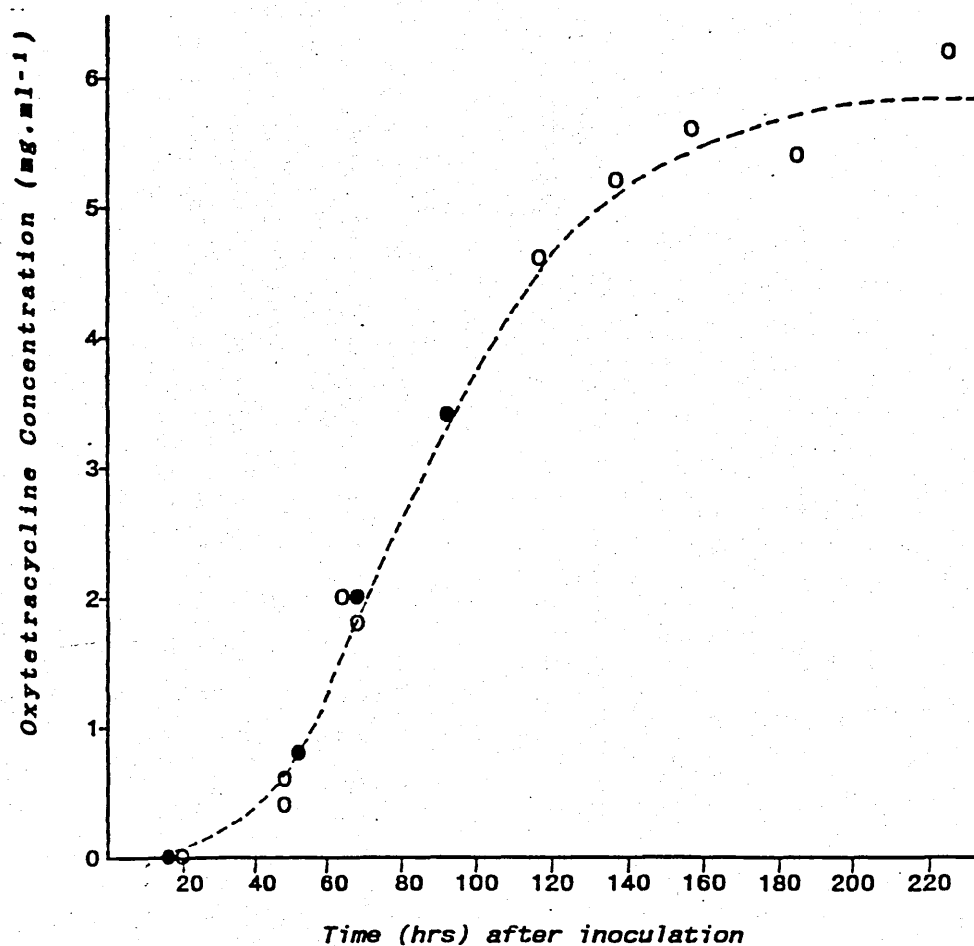
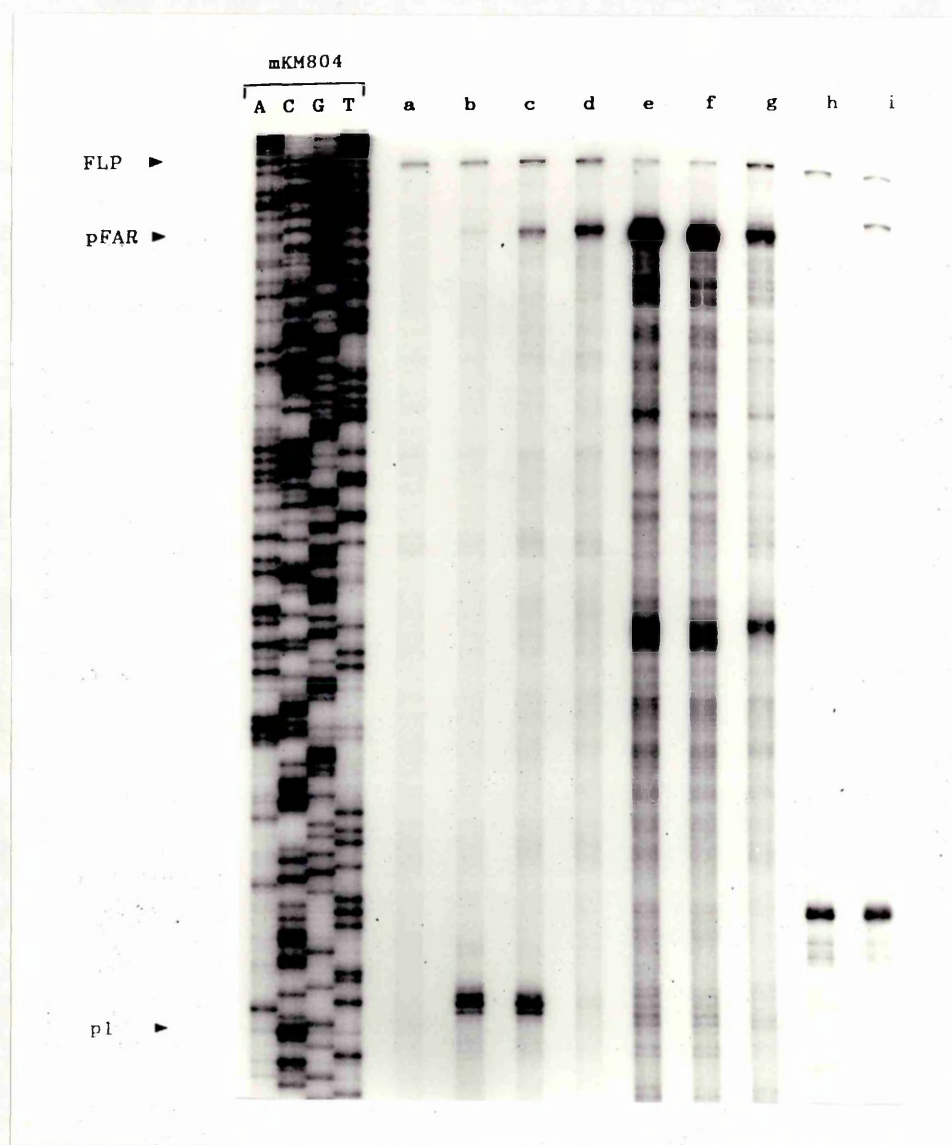


Figure 3.6. Oxytetracycline production by *S. rimosus* M15883 grown in Trusoya Medium 1. The plot is derived from 14 samples (O) taken from different (50ml) cultures, inoculated at the same time and incubated under standardised conditions. The (●) symbol indicates the concentration of OTC in cultures from which RNA was isolated for S1 nuclease protection experiments (see Figure 3.7).



**Figure 3.7. (A)** S1 nuclease protection by transcripts of the *otrA* resistance gene from *S. rimosus* M15883 grown on TS1. Single-stranded probe ( $2.9 \times 10^7$  Cerenkov cpm.pmol<sup>-1</sup>), was generated by extension from universal primer, which was annealed to mKM804, and digestion with *EcoRI* (Materials and Methods, 2.3.18). 0.2 pmoles of probe was hybridized at 52°C with 10 µg of RNA from the following: (a) *S. rimosus* M15883S grown on TSB medium; (b) *S. rimosus* M4018 grown on TSB (prepared by Doyle, 1987); (c) *S. rimosus* M4018 grown on TSB medium (this work); (d); (e); (f) and (g) *S. rimosus* M15883 grown on TS1 (the OTC production medium) for 18, 46, 69, and 93 hours respectively; (h) and (i) as (b) and (c) respectively. Samples (a) to (g) were digested with S1 nuclease. Exonuclease VII was used to digest samples (h) and (i). The positions in the mKM804 sequencing ladder that are coincident with the RNA-protected fragments derived from transcripts originating from *otrA*p1 and far upstream are indicated by p1 and pFAR. The position of probe protected along its entire length by contaminating template DNA (FLP) is also indicated.

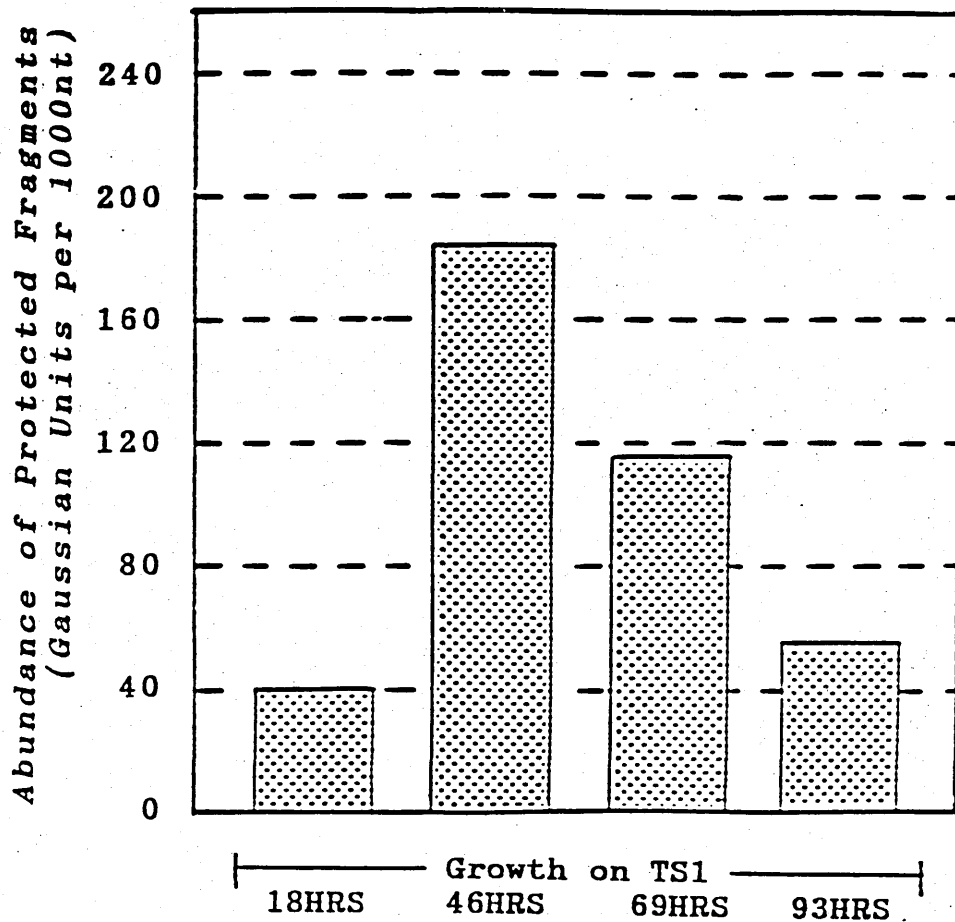


Figure 3.7. (B) Histogram showing the results of quantitative analysis of the *otrA* transcripts originating from p1 and far upstream. Abundance estimates were obtained by densitometry scanning (Materials and Methods, 2.3.21) of autoradiographs produced by different exposures to the denaturing polyacrylamide gel in which the RNA-protected fragments had been resolved (panel A).

far upstream at the onset of antibiotic production. The abundance of *otrA* transcripts from pFAR within the total RNA population increased by a factor of 4, between 22 and 46 hours. Over the next two time points, 69 and 93 hours, the level of these transcripts declined steadily to almost the 22 hour level. This pattern of transcription from pFAR has been reproduced in a separate S1-nuclease protection experiment (result not shown). RNA could not be isolated from cultures older than 100 hours, suggesting that the majority of protein synthesis and cellular metabolism had ceased by this point in the fermentation.

Exonuclease VII digestion of hybridization complexes formed between the *HincII*-*Bam*HI probe and RNA from *S. rimosus* M4018 (Figure 3.7, lanes b and c) produced RNA-protected fragments of similar size to those produced by S1 nuclease digestion (Figure 3.7, lanes h and i). This indicated that the proposed transcription start sites for *otrA*p1 at 200-221nt (Doyle, 1987) were not an artifact of endonucleolytic cleavage at a site of secondary structure caused by a breakdown in complementarity between the *otrA* message isolated from *S. rimosus* M4018 and the DNA probe derived from *S. rimosus* M15883. Exonuclease VII can only digest unprotected probe from the 3' and 5' ends to within 5-7nt of the DNA-RNA hybrid (Rose and Botstein, 1983). Consequently, the Exonuclease VII-resistant fragments (lanes h and i) are 10-14nt longer than the corresponding S1 nuclease-resistant fragments (lanes b and c).

### 3.3 Discussion

The nuclease protection experiments and *in vivo* promoter analysis described in this chapter, have shown clearly that the *otrA* resistance gene is transcribed from *otrA*p1 and a promoter(s) located far upstream, the former resembling the prokaryotic consensus sequence (Doyle, 1987). It is now believed that overdigestion of a DNA-probe, which was protected by transcripts originating from far upstream of *otrA*, with S1 nuclease led previously to the erroneous identification of a promoter that overlapped the *Bam*HI site (Doyle, 1987).

Interestingly, the maximum level of kanamycin resistance conferred by pKMS605 (containing the 305bp *Bcl*I5-*Kpn*I7 region) is at least three fold greater than that conferred by pDM130 (containing the *Sma*I4-*Bam*HI6 region). The pIJ486-construct pDM110, that contains the *Bcl*I5-*Bam*HI6 region, conferred the same level of resistance as pDM130 (Doyle, 1987). The increased kanamycin levels specified by pKMS605 appeared to be coincident with the inclusion of the 91bp *Bam*HI6-*Kpn*I7 region in the cloned DNA.

Tetracycline resistance in *S. rimosus* ATCC 10970 and *S. rimosus* M4018 can be induced by pre-exposure to sub-lethal concentrations of tetracycline (Ohnuki *et al.*, 1985; Lee, unpublished results). The *otrA* homologue from *S. rimosus* ATCC 10970, designated *tetA*, has been cloned in *Streptomyces griseus* using a plasmid vector. The demonstration that expression of *tetA* was also inducible in this heterologous host suggested that; (i) any DNA segment(s) required for the regulation of *tetA* expression was also cloned in the plasmid, (ii) *S. griseus* contained the necessary factor(s) required for the induction of *tetA* expression and (iii) that the induction mechanism(s) was not unique to *S. rimosus* or oxytetracycline biosynthesis. The induction of high levels of tetracycline resistance conferred by *tetA* has not been investigated further.

In *E. coli* and *Salmonella typhimurium* the transposon Tn10 carries two genes, *tetA* and *tetR*, which confer high-level, inducible resistance to tetracycline (Kaneko *et al.*, 1985). The amino acid sequence of the deduced *tetA* protein has high identity to the predicted amino acid sequence of the *tetB* (*otrB*) gene product from *S. rimosus* (Reynes *et al.*, 1988). The protein products of both *tetA* of Tn10 and *tetB* of *S. rimosus* have been shown to mediate the efflux of tetracycline from the cell (Kaneko *et al.*, 1985; Ohnuki *et al.*, 1985). The *tetR* gene encodes a repressor protein which regulates negatively transcription of both the *tetA* and *tetR* genes, which are transcribed from divergent, overlapping promoters. Tetracycline binds to the *tetR* repressor reducing its affinity for the *tet* operator and thus, induces expression of both *tetA* and *tetR* (Hillen *et al.*, 1984). Sequencing of the *act* cluster of *S. coelicolor*

within a region that confers resistance to actinorhodin has identified PPCRs, which are predicted to encode proteins with high amino acid sequence identity to the Tn10 *tetA* and *tetR* products (Hopwood, pers. comm.). It is possible that actinorhodin acts as an inducer of resistance in *S. coelicolor*.

Sequencing of the *otc* cluster of *S. rimosus* around the resistance determinants has not identified any PPCR which is predicted to encode a regulatory protein that is homologous to the *tetR* protein from Tn10. It is unlikely, therefore, that the DNA fragments from *S. rimosus*, which confer tetracycline resistance in *S. griseus* (Ohnuki *et al.*, 1985), encode a regulatory gene similar to *tetR* of Tn10. The mechanism by which tetracycline resistance in *S. griseus* can be induced remains to be elucidated. *S. griseus* may encode a regulatory protein, which is similar to *tetR* and binds to a regulatory sequence within the promoter region of *tetA* (*otrA*) from *S. rimosus* and thus blocks transcription. However, this proposal raises a number of questions. What are the host-encoded gene(s) that the putative regulator protein would regulate normally? How do sub-lethal concentrations of tetracycline induce tetracycline resistance? Does tetracycline bind the putative regulatory protein, decreasing its affinity for an operator? *S. griseus* is not known to produce any secondary metabolites, which are related structurally to tetracycline. What is the structure of the inducer? Can sub-lethal concentrations of other antibiotics induce tetracycline resistance in transformants of *S. griseus* containing resistance determinants from *S. rimosus*? Although there are precedents for repressor/induction mechanism in regulating tetracycline resistance in Tn10 and possibly actinorhodin resistance in *S. coelicolor*, it is possible that the inducible tetracycline resistance conferred by *S. rimosus* determinants in *S. griseus* is mediated by an alternative, unknown mechanism. Analysis of the *otrAp1* promoter region has not identified any sequence, which is similar to the *tet* operator in Tn10 (Hillen *et al.*, 1986).

As the minimum length of cloned DNA that could confer inducible tetracycline resistance was 2.5kb (Ohnuki *et al.*, 1985) and the protein coding sequence of *otrA* extends for 2kb (Doyle, 1987), any regulatory sequences required for inducible expression are probably contained within 500bp upstream of translation start codon of *tetA*. Assuming that the transcriptional organisation and regulation of OTC biosynthesis are similar in *S. rimosus* M15883 and ATCC 10970, it is possible that the pKMS605 construct, which contains sequence extending from *otrA*p1 to 435bp upstream of the protein coding region sequence of *otrA*, contains any regulatory sequences required for the inducible expression of *otrA*. This could be tested easily by introducing pKMS605 into *S. griseus* and determining if kanamycin resistance, which is conferred by the *aphII* reporter gene, can be induced by pre-exposure to sub-lethal concentration of tetracycline (and/ or oxytetracycline). It would also be interesting to determine if expression of the *aphII* gene can also be induced by tetracycline in the *S. lividans* transformants containing pKMS605.

Although the induction of expression of *otrA* or *tetA* has not been investigated further, the significantly lower kanamycin resistance levels conferred by pDM130 and pDM110 compared with pKMS605 could result from the disruption of the mechanism(s) which regulates expression of *otrA*. As there are 220bp between the transcription start sites of *otrA*p1 and the *Bam*HI<sub>6</sub> site, sequence(s) distant from this promoter could be implicated in the regulation of *otrA* expression. Although, the binding of regulatory protein(s) to recognition sequences on DNA far from the point of transcription initiation was discovered initially in eukaryotic cells, similar regulatory interactions have been found in prokaryotes (for an overview see Magasanik and Neidhardt, 1987; Magasanik, 1989) and may be important in regulating the transcription from *otrA*p1. It is envisaged generally that DNA looping (Ptashne, 1986) facilitates proteins bound at a distance from the promoter sequence to interact either with other regulatory proteins bound near to the promoter or with the RNA polymerase directly. For example, the repression of transcription of the *lacZYA* operon in *E. coli* requires the cooperative binding of the *lacI* repressor to two operator sites (for review see

Gralla, 1989 ). The primary operator  $O_1$  overlaps with the promoter sequence (Gilbert *et al.*, 1976), while the secondary operator  $O_2$  is located within the *lacZ* gene, 401bp downstream of the transcription start point (Reznikoff *et al.*, 1974). The *lac* repressor has been shown to bind simultaneously to  $O_1$  and  $O_2$ , causing the intervening DNA to be looped out (Kramer *et al.*, 1988). The cooperative binding of the repressor to both of these operator sites is thought to repress transcription of *lacZ* in two ways; (i) by tightening the interaction of the repressor bound at  $O_1$ , thereby, preventing RNA polymerase from binding to the promoter and (ii) by preventing the transcription of *lacZ* proceeding beyond the secondary operator ( $O_2$ ), within the  $\beta$ -galactosidase gene (Flashner and Gralla, 1988). An analogous arrangement of operator sites has also been found in the *ara* operon (Huo *et al.*, 1988; Hamilton and Lee, 1988) and the *gal* operon (Irani *et al.*, 1983; Haber and Adhya, 1988). The mechanisms which have been proposed to regulate transcription of the *ara* and *gal* operons also involve the cooperative binding of repressor protein and the looping out of DNA. Should transcription from *otrAp1* be subject to repression (although there is no experimental evidence to distinguish between an activation or repression mechanism), it is difficult to envisage how the addition of an upstream region would lead to "derepression".

The disparity in the level of resistance to kanamycin conferred by pDM130 and pKMS605 could be explained should maximum transcription from *otrAp1* require an interaction with a protein bound to DNA far upstream (at least 200bp) of the transcription start point. In prokaryotes, only a single example of a pathway-specific activator protein, which binds at a distance from promoter sequences, has been described. The promoter of the *glnALG* operon is controlled by the nitrogen regulator  $NR_1$  (also called NtrC or GlnG), which is activated by phosphorylation in response to nitrogen starvation (for review see Magasanik, 1988). The looping of DNA is thought to facilitate the activation of RNA polymerase holoenzyme  $E\sigma^{54}$ , which is prebound to the promoter, by facilitating an interaction with phosphorylated  $NR_1$  (P- $NR_1$ ) bound about 100bp upstream of the transcription start point (Sasse-Dwight and Gralla, 1988). The binding site for P- $NR_1$  can be moved more



than a kilobase away from the promoter and still facilitate transcriptional activation (Reitzer and Magasanik, 1986). A similar activation mechanism has also been identified in other systems regulated by nitrogen, for example, the transcriptional activation of the formate dehydrogenase gene in *E. coli* (Gralla, 1990). It has been proposed that Eo<sup>54</sup> is a specific receptor for the NR<sub>1</sub> regulator (Gralla, 1989 ).

The analysis of transcriptional activation by the OmpR protein has revealed that activation can be mediated from locations several turns of the DNA helix away so long as the binding site remains on the original side of the DNA (Maeda *et al.*, 1988). However, the ability to activate transcription diminished significantly with distance. It was suggested that activator proteins such as CRP and OmpR, which activate Eo<sup>70</sup>, may be unable to act over long distances in *E. coli* (Gralla *et al.*, 1989 ). To date, the transcriptional activation of the stable RNA (ribosomal and transfer RNA) operons of *E. coli* by the binding of FIS to DNA determinants up to 130bp upstream of the transcription start sites (Nilsson *et al.*, 1990) represents the furthest distance that a protein bound to DNA has been shown to activate transcription by Eo<sup>70</sup>.

In *Bacillus subtilis*, hyperprotease production (*hpr*) mutants have been isolated, which result in increased expression of the neutral protease (*nprE*) and alkaline protease (*aprE*) genes (Henner *et al.*, 1988). Northern analysis has shown that the *hpr* mutations affect the expression of the *aprE* gene at the level of transcription (Ferrari *et al.*, 1988). Deletion analysis of *aprE'*-*'lacZ* fusions in *hpr* mutants has revealed that stimulation by *hpr* is lost somewhere between 200 and 400bp upstream of the transcription initiation site. Although the mechanism by which *aprE* expression is activated remains to be elucidated, this example nevertheless demonstrates the importance of DNA determinants located "far upstream" in regulating gene expression. Therefore, it is a formal possibility that maximum expression of *otrA* requires regulatory sequences located within the *Bam*HI6-*Kpn*I7 region, at least 220bp upstream of the transcription start point of *otrA*p1.

The differences in kanamycin levels between pDM110 and pKMS605 could also reflect differences in the topology of the DNA in the immediate vicinity of *otrAp1*. It is well established that promoter function *in vivo* and *in vitro* can be influenced strongly by the level of DNA supercoiling (for review, see Drlica, 1987). For example, the specific induction of transcription of the *proU* loci of *Escherichia coli* and *Salmonella typhimurium* is mediated by changes in the DNA superhelicity *in vivo* in response to osmotic stress (Higgins *et al.*, 1988). Analysis of the *leu-500* promoter of *S. typhimurium* (Richardson *et al.*, 1988) has shown that local rather than global changes in DNA topology can be important in activating transcription. Whilst changes in the topology of DNA will clearly be influenced by changes in the activities of topoisomerase I and gyrase, the local topology may also be dependent on the binding of proteins such as HU and H1 (the general histone-like DNA-binding proteins) or more specifically by factors similar to FIS (Factor for Inversion Stimulation: described by Kahman *et al.*, 1985; Johnson and Simon, 1985) or IHF (Integration Host Factor: described by Craig and Nash, 1984).

Although a number of inverted repeats and direct repeats, which could be the binding sites for regulatory proteins, are present upstream of *otrAp1* in the region inserted in the pKMS605 construct (result not shown), no sequence similarity has so far been detected to any consensus binding sites for DNA-binding proteins. It would not be constructive to discuss the repeats located upstream of *otrA* in any detail, as streptomycetes have an inherent capacity to form such structures at high frequency, due to the high G+C composition of the DNA and the restricted use of synonymous codons.

The above is not intended to be an exhaustive discussion of the mechanisms which could account for the disparity in the levels of kanamycin resistance conferred by pKMS605 and pDM130. It is intended merely to demonstrate the scope of possible regulatory interactions. As the first step in analysing the tentative regulation of transcription, a series of promoter-probe constructs containing *otrAp1*, with the upstream region truncated to different extents, is currently being made to delineate the sequence(s) required for the maximal activity of *otrAp1 in vivo* (McDowall, unpublished work).

Using high-resolution S1 nuclease protection experiments to analyse RNA samples isolated from cultures at different stages in the fermentation of oxytetracycline, it is possible to quantitate the relative changes in the abundance of transcripts from *otrA*<sub>pl</sub> and far upstream. However, only general inferences can be made about the absolute levels of *otrA* transcripts within the mycelium, as the rates of accumulation and turnover of rRNA, tRNA and mRNA in streptomycetes have not been studied in any detail. Physiological studies using *Escherichia coli* as a model have indicated that the rate of synthesis of rRNA and tRNA is regulated and corresponds to the cellular need for ribosomes (for an overview see Jinks-Robertson and Nomura, 1987). As the growth rate of *E. coli* slows, the demand for protein synthesis decreases and there is a corresponding decrease in the number of ribosomes per cell. As rRNA and tRNA account for the majority of RNA in *E. coli* (greater than 95% in cells growing exponentially) a slowing of the growth rate results in an overall reduction in the total RNA per cell. The growth rate dependence of ribosome accumulation is almost certainly a ubiquitous feature of bacterial growth, due to the central role of ribosomes in the synthesis of other cellular components and the large amounts of cellular energy and metabolic intermediates required to produce ribosomes.

A striking reduction in protein, DNA and RNA synthesis has been detected in the transitory phase between vegetative growth and morphological development in *Streptomyces coelicolor* A3(2), when grown on solid medium (Granozzi *et al.*, 1990). This reduction in macromolecular synthesis is similar to the change seen in *E. coli* cultures entering stationary phase. In *S. coelicolor* A3(2), however, the synthesis of macromolecules resumes at the onset of morphological development and is restored nearly to the values measured during vegetative growth (Granozzi *et al.*, 1990). The production of secondary metabolites coincides typically with morphological development. Indeed, substantial genetic evidence suggests that both processes share elements of a complex regulatory network and at least in principle could be triggered by a common signal (General Introduction, 1.9). The accumulation of RNA during the fermentation of oxytetracycline may also occur in two distinct phases.

Analysis of the transcription of *otrA*, using RNA isolated from *S. rimosus* M4018 grown on Liquid Complete Medium, has shown that this resistance gene is transcribed almost exclusively from *otrAp1* during intense vegetative growth. The relative contribution of *otrAp1* diminishes, however, as the growth rate slows and the onset of oxytetracycline production approaches. Early in the production phase, transcripts originating from *otrAp1* and far upstream are almost equal in abundance.

The changes in the relative abundance of transcripts of *otrA* were seen against an overall decrease in the abundance of these transcripts within the total RNA population. Any change in the overall RNA content of the *S. rimosus* M4018 mycelium, as the culture approaches stationary phase and the onset of antibiotic production, is likely to be a reduction in the total amount of rRNA, tRNA and mRNA (associated with vegetative growth). The decrease in the abundance of *otrA* transcripts, within the total RNA population at this early stage in the fermentation, therefore, probably represents a real decrease in the actual number of transcripts within the mycelium. However, before credible predictions can be made about the absolute levels of transcripts, it will be necessary to study the rates of accumulation and turnover of mRNA, rRNA and tRNA at different stages in the life cycle of *Streptomyces*.

The analysis of transcription during the production phase of a fermentation failed to detect transcription from *otrAp1* in RNA isolated from *S. rimosus* M15883 grown on TS1. Transcripts originating from far upstream increased in abundance during the early stages of production (relative to total RNA), peaking at about 20 hours after oxytetracycline was first detected in the culture broth, before decreasing in abundance to levels seen at the onset of production. Should a second round of RNA synthesis occur at the onset of oxytetracycline production in liquid cultures of *S. rimosus*, as described for the morphological development of *S. coelicolor* A3(2) on solid medium (Granozzi *et al.*, 1990), then the transient increase in the abundance of transcripts from far upstream, within the total RNA population, may represent an actual increase in

their numbers within the mycelium. It is proposed that the transient peak in the abundance of *otrA* transcripts from far upstream within the total RNA population probably reflects transient activity at the cognate promoter(s), located far upstream.

A similar observation has been reported recently by Guthrie and Chater (1990). The transcripts of an undescribed gene in the undecylprodigiosin cluster of *S. coelicolor* increased briefly in abundance after the most rapid growth phase had passed (Guthrie and Chater, 1990). The level of transcription in the undecylprodigiosin cluster was estimated by the insertion of a promoterless *xylE* gene (described in Chapter 5) into a previously undescribed region implicated in the early steps of biosynthesis, using a bacteriophage-based integrative vector.

The results presented in this chapter indicate that the expression of the *otrA* resistance gene is under complex transcriptional control. The two transcripts of *otrA* which have been identified are transcribed from promoters with distinct temporal activities. In *Bacillus subtilis* a cascade of alternative sigma factors enables RNA polymerase to recognise unique classes of promoters and thereby activate selective sets of sporulation genes, in the correct order and at the appropriate time (Losick and Pero, 1981). The finding that the switch from continued extension of aerial hyphae to their development into chains of spores, is controlled by an alternative sigma factor ( $\sigma^{whiG}$ ) in *S. coelicolor*, suggests that a similar regulatory cascade could be important in controlling the differentiation of *Streptomyces* (Chater *et al.*, 1989).

The putative -35 region (CTGACC) and -10 region (TACCGT) of *otrA*<sub>p1</sub> resembles the -35 region (TTGACA) and -10 region (TATAAT) of the consensus for the major class of eubacterial promoters (Hawley and McClure, 1983). Correspondingly, this promoter has a distinct temporal activity associated with vegetative growth. Using the pK0500 promoter-probe vector (Close and Rodriguez, 1982), the *Bam*HI<sub>5</sub>-*Bcl*I<sub>6</sub> fragment containing the *otrA*<sub>p1</sub> promoter was shown to be capable of directing transcription in *E. coli* (McDowall, unpublished result). Therefore, it

would appear that the *otrAp1* promoter is recognised by a form of RNA polymerase in *Streptomyces* which is responsible for the transcription of the general housekeeping genes.

Comparative analysis of promoters responsible for the transcription of genes involved in morphological and physiological differentiation could define common elements involved in the control of *Streptomyces* development. Therefore, the promoter responsible for transcription of *otrA* from far upstream was isolated and characterised as the next step in dissecting the coordination of oxytetracycline resistance and production (see Chapter 5).

## **CHAPTER 4**

### **THE *OTCZ* GENE: NUCLEOTIDE SEQUENCE, DEDUCED FUNCTION AND TRANSCRIPTION**

## 4.1 Introduction

The *otc-151* mutant is representative of the *otcZ* class of mutants that are blocked in "early" oxytetracycline (OTC) biosynthesis, before the formation of anhydrotetracycline (ATC). The specific enzymatic step which is blocked has not been assigned, as the *otcZ151* mutant failed to produce OTC in cosynthesis tests when it was combined with representatives from other groups of "early" mutants on agar strips and was also unable to convert purified "early" intermediates to OTC when these were supplied to liquid cultures (Rhodes *et al.*, 1981).

Antibiotic production was restored in the *otcZ151* mutant by the introduction of a plasmid containing the 4.4kb *SphI*<sub>3</sub>-*SphI*<sub>12</sub> region (Figure 4.1) (Butler *et al.*, 1989). The mutation on the chromosome was either complemented *in trans* by the expression of a functional gene from the plasmid or it was repaired by a recombination event(s) between homologous sequences on the chromosome and DNA cloned on the plasmid. Using a series of plasmids that contained truncated derivatives of an insert, which in full-length form could restore antibiotic production in *otcZ151*, the mutation was mapped close to the "left" of the *PstI*<sub>8</sub> site (Butler *et al.*, 1989).

The extreme "left" of the 4.4kb *SphI* region (*SphI*<sub>3</sub>-*BclI*<sub>5</sub>) contains the 5' end of the *otrA* resistance gene, which has been sequenced (Doyle, 1987). The *otcC* gene, encoding the ATC oxygenase, has also been localised within the *SphI*<sub>3</sub>-*SphI*<sub>12</sub> region using "reverse genetics" (Binnie *et al.*, 1989). Purification of the ATC oxygenase allowed "reverse translation" of the amino acid sequence from the amino-terminus. An oligonucleotide containing the predicted nucleotide sequence was synthesized chemically and used as a probe to locate the 5' end of the *otcC* gene to the "left" of the *SphI*<sub>12</sub> site. Limited sequencing suggested that the potential protein coding region (PPCR) of *otcC* read from "right" to "left". From the molecular weight of the purified ATC oxygenase (52 kD), the 3' end of the *otcC* gene was predicted to be between the restriction sites *SmaI*<sub>9</sub> and *PstI*<sub>8</sub>.



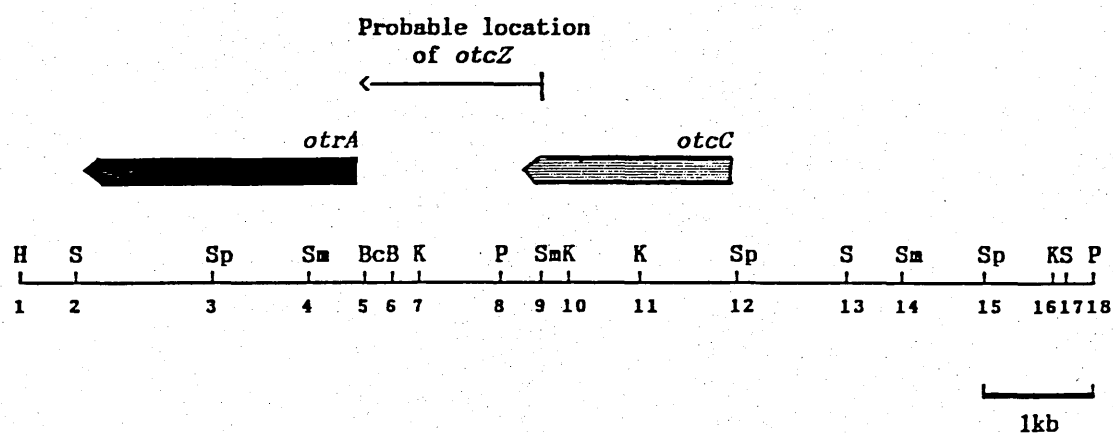


Figure 4.1. Restriction map of the "left" end of the oxytetracycline cluster. The location of *otcA* and the probable locations of *otcC* and *otcZ* are indicated. For clarity only the restriction sites that are relevant to the delineation of the position of *otcZ* are shown. B, *Bam*HI; Bc, *Bcl*I; H, *Hind*III; K, *Kpn*I; S, *Sst*I; Sm, *Sma*I; Sp, *Sph*I; P, *Pst*I.

The above results delineated the possible location of the DNA segment, which "complements" the *otcZ151* mutation, to the 1.3kb *BcII*-*SmaI* region. The determination of the DNA sequence in this region was considered a priority for the following reasons.

(i). The DNA sequence could be used to predict the precise genetic organisation of this region and the primary structure of the *otcZ* gene product.

(ii). The inferred protein sequence could then be compared with entries in computer-held databases, such as GenBank, EMBL and NBRF, for similarity to proteins of known biological function. Homology with other proteins might provide insight into the enzymatic step blocked in the *otcZ151* mutant.

(iii). A knowledge of the genetic organisation, in conjunction with an analysis of transcription, would be necessary to identify regulatory sequences involved in controlling the expression of the oxytetracycline genes.

## 4.2 Results

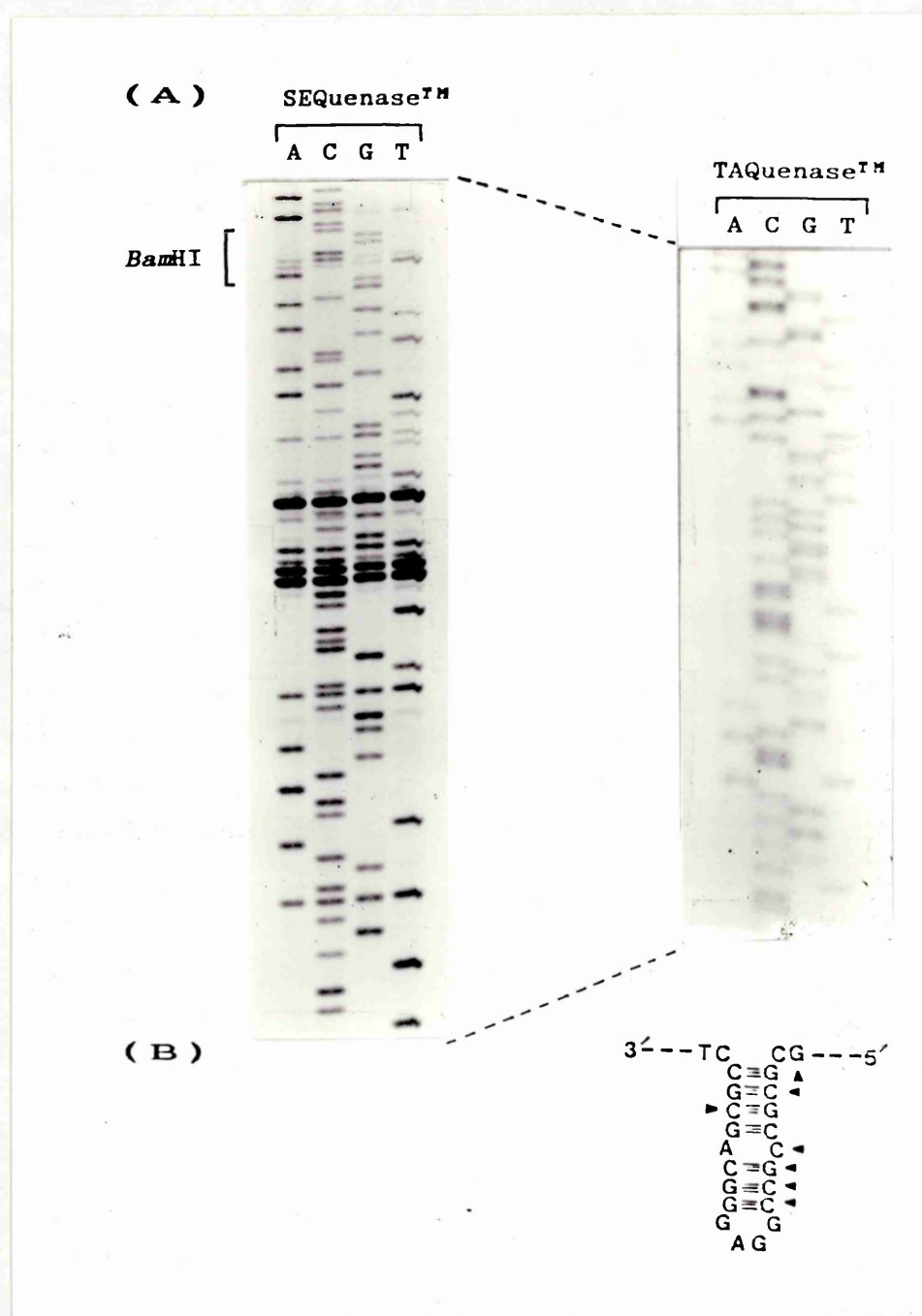
### 4.2.1 Sequencing Method

Modified DNA polymerase from *Thermus aquaticus* (Taq polymerase) was used to sequence the *otcZ* gene. This thermo-stable enzyme has a temperature optimum of 80°C, which is considerably higher than that of other commercially-available DNA polymerases for dideoxy sequencing, such as modified T7 DNA polymerase (supplied by United States Biochemical Corporation as SEQuenase™) and modified DNA polymerase I from *Escherichia coli* (supplied by BRL as Klenow), which are used at recommended incubation temperatures of 42°C and 30°C respectively. The high incubation temperature of Taq polymerase was found to be a considerable advantage when sequencing DNA from streptomycetes, which typically has a base composition of 73% mol G+C (Enquist and Bradley, 1971). Inverted repeat sequences occur frequently in *Streptomyces* DNA and these can hybridize to form highly stable G+C-rich secondary structures.

The effects of secondary structure can be manifested at two different stages during sequencing. Firstly, stable hairpin loops in the template can cause the DNA polymerase to "stall" during the sequencing reaction and as a result artifactual bands are produced in all four tracks of the sequencing ladder. Secondly, the products of the sequencing reaction may form regions of G+C-rich secondary structure. This can cause bands to be compressed, and result in part of the sequencing ladder being unreadable over a localised area.

Dideoxy sequencing of DNA from *Streptomyces* has been undertaken previously at Glasgow using Klenow DNA polymerase (Doyle, 1987). Secondary structure posed a significant problem and it was necessary to use elevated incubation temperatures, formamide gels and the chemical degradation method of Maxam and Gilbert (1977) to complete the sequence of the *otrA* resistance gene. A number of alternative polymerases for DNA sequencing have become readily available since the sequencing of *otrA*. These include SEQuenase<sup>TM</sup>II and TAQuenase<sup>TM</sup> (modified Taq polymerase), both supplied by USB.

Before a strategy for the sequencing of *otcZ* was devised, the ability of these polymerases to produce readable sequence from a single-stranded template (mKM905) known to cause the Klenow enzyme to "stall" during sequencing was investigated using the standard protocols supplied with the USB kits (Figure 4.2). The SEQuenase<sup>TM</sup> enzyme apparently had a considerable problem reading through regions of streptomycete DNA at the recommended incubation temperature of 42°C (Figure 4.2, Panel A). At the higher incubation temperature used with Taq polymerase (70°C) the G+C-rich secondary structures (Figure 4.2(B)) were presumably less stable and consequently the progression of the sequencing reaction was not impeded. The TAQuenase<sup>TM</sup> sequencing kit was used, therefore, for the sequencing of *otcZ*. Termination mixes containing the nucleotide analogue 7-deaza dGTP were used routinely to minimise the possibility of band compressions in the sequencing ladders. This analogue pairs weakly with conventional bases and thus, reduces the possibility of secondary structure formation in the products of the sequencing reaction (Mizusawa *et al.*, 1986).



**Figure 4.2.** (A) Comparison of the sequencing reactions using SEQuenase™ and TAQuenase™. (B) Secondary structure predicted by STEMLOOP in the region of the premature terminations. Small arrows (►) indicate the precise positions at which the SEQuenase enzyme stalls on the template. For the SEQuenase™ reactions, the radio-label was [<sup>35</sup>S] α-dATP. [<sup>32</sup>P] α-dATP was used for the TAQuenase™ reactions.

It has been found subsequently that other groups have obtained good quality sequence data from streptomyces DNA using SEQuenase™, when the reactions are carried out at elevated incubation temperatures (ca. 55°C) and used in conjunction with single-stranded DNA-binding protein from bacteriophage T4 (Leadlay, pers. comm.).

#### 4.2.2 Sequencing Strategy

A summary of the sequencing strategy is shown in Figure 4.3. Three internal fragments from the 1.3kb *Bcl*I<sub>5</sub>-*Sma*I<sub>9</sub> region were ligated to M13mp18 and M13mp19 respectively, which had been digested with the appropriate restriction endonucleases, and introduced into *E. coli* TG1 by genetic transformation (Section 2.3.12). The single-stranded templates were isolated from phages containing recombinant DNA molecules (Section 2.3.16). The 250bp *Pst*I<sub>8</sub>-*Sma*I<sub>9</sub> fragment from pKME917 (pUC19 containing the *Sma*I<sub>4-9</sub> fragment) was inserted into M13mp18 and M13mp19 to generate mKM803 and mKM903 respectively. The 722bp *Kpn*I<sub>7</sub>-*Pst*I<sub>8</sub> and the 305bp *Bcl*I<sub>5</sub>-*Kpn*I<sub>7</sub> fragments were obtained from the pUC18-based constructs, which were intermediates in the introduction of the same fragments into the pIJ486 promoter-probe vector (Section 3.2.1). *Hind*III/*Eco*RI fragments from these constructs were then inserted into M13mp18 and M13mp19 to derive mKM807 and mKM907 from pKME807 and mKM805 and mKM905 from pKME805 respectively.

The 722bp *Kpn*I-*Pst*I inserts in the mKM807 and mKM907 templates could not be sequenced completely using the universal primer (UP), as only around 300 nucleotides could be read from the standard sequencing reactions. Oligonucleotides 807a and 907a, which incorporated sequence obtained from extension of the universal primer were used, therefore, as primers to extend the sequencing of the mKM807 and mKM907 templates respectively. In turn, sequence data generated from the 807a and 907a primers was used to design oligonucleotides 807b and 907b respectively, which allowed the mKM807 and mKM907 templates to be sequenced completely.

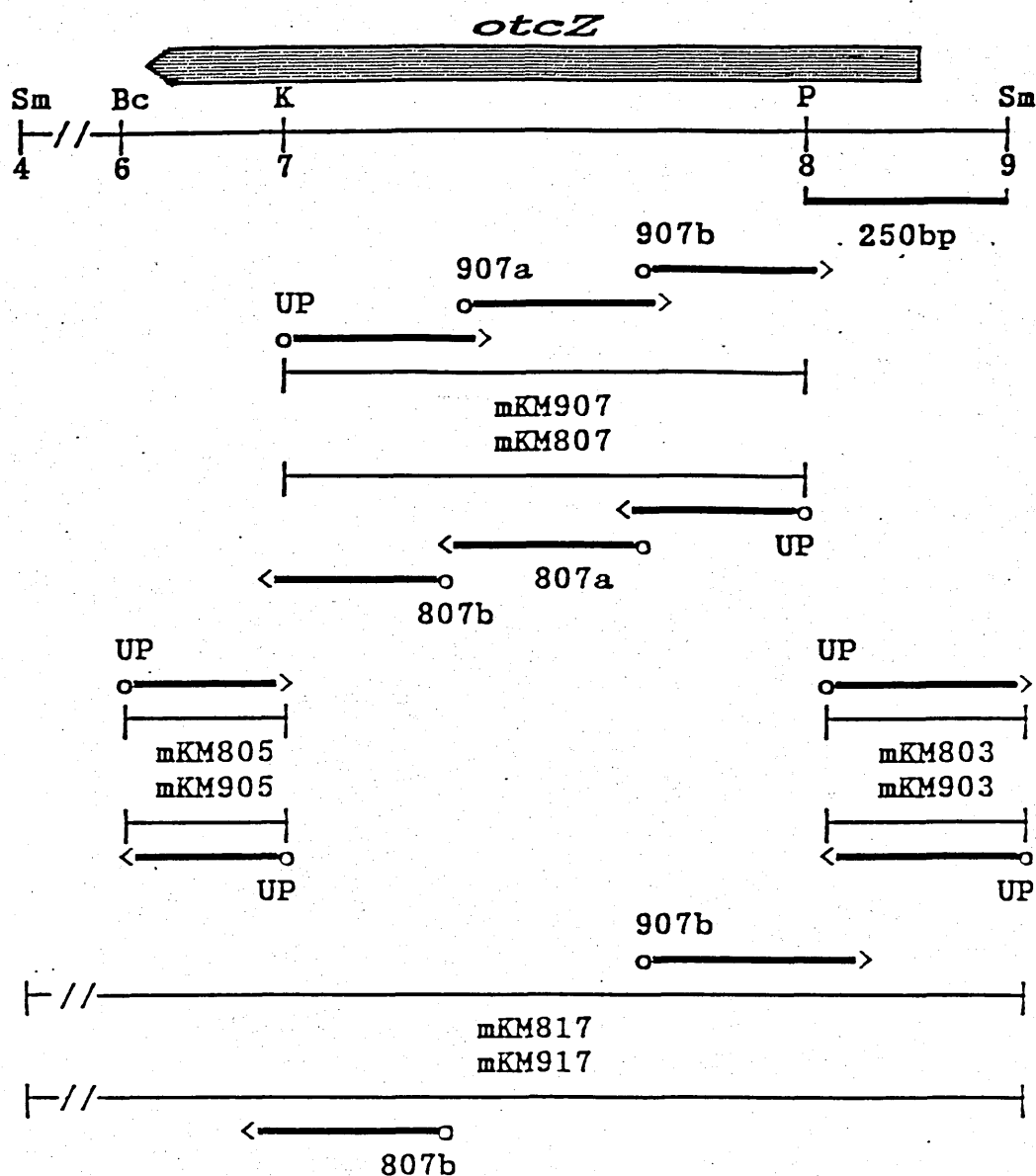
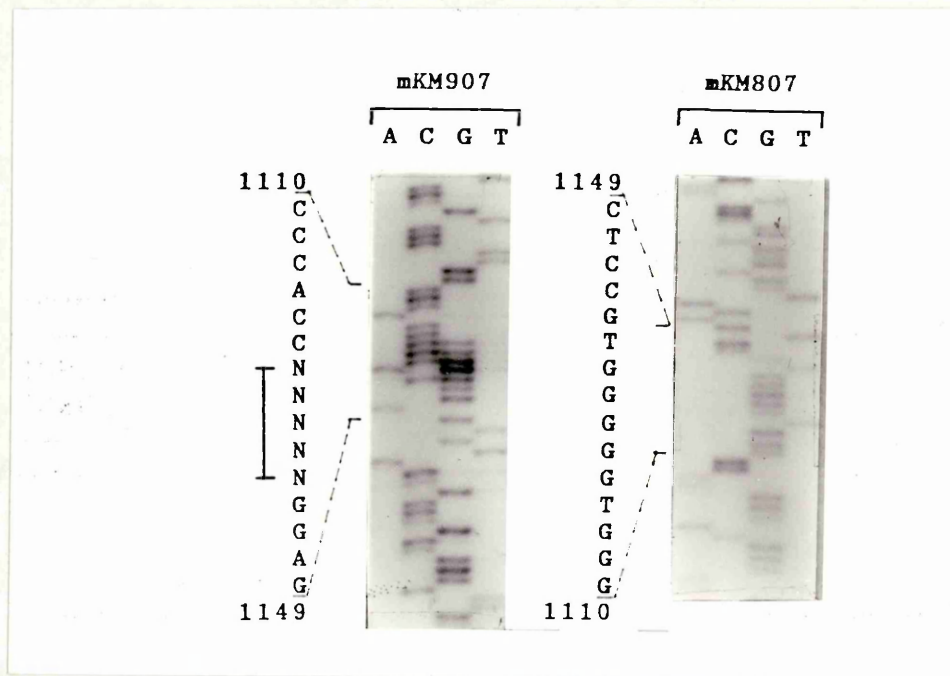


Figure 4.3. M13 templates used to sequence the 1.3kb *BcII*-*SmaI* region. The template designations are given above the lines (|---|) used to illustrate the extent of the cloned DNA. The arrows indicate the direction and extent of primed synthesis during nucleotide sequence analysis. The oligonucleotides used to prime the extension reactions are given above the arrows. UP, represents the universal primer. For clarity, only restriction sites relevant to the subcloning of the *BcII*-*SmaI* region are shown.



**Figure 4.4.** Comparison of sequence ladders produced from mKM907 and the complementary template mKM807. A line (|) indicates the extent of the sequencing artifact seen in the mKM907 ladder but not evident at the corresponding position in the mKM807 ladder.

Only a single stretch of 5 to 6 nucleotides from the mKM903 template could not be read unambiguously due to the comigration of bands from different tracks in the sequencing ladder (Figure 4.4). Interestingly, this region (1110-1130nt) is not predicted to form significant secondary structure. The sequence data from the complementary strand (mKM803) was excellent and no attempts were made to overcome this artifact. Upon completion of this work, non-overlapping regions of sequence remained at the *KpnI*<sub>7</sub> and *PstI*<sub>8</sub> boundaries. Two other M13mp18-derived constructs were used to sequence across these restriction sites. The mKM817 construct contained the 1.6kb *SmaI*<sub>4</sub>-*SmaI*<sub>9</sub> fragment, which was subcloned from pKME917 as a *HindIII*/*EcoRI* fragment, with the *SmaI*<sub>9</sub> site nearest the *EcoRI* site in the polylinker. The mKM971 construct contained the same insert but in the opposite orientation. The mKM817 template, in conjunction with the oligo. 807b, was used to sequence across the *KpnI* boundary. The *PstI* boundary was sequenced using mKM917 and oligo. 907b, thereby completing the sequencing of the 1.3kb *BclI*<sub>5</sub>-*SmaI*<sub>9</sub> region.

The sequence is comprised of 1,265bp and has a base composition of 72.57 mol% G+C. The annotated sequence, including the amino acid sequences of potential protein coding regions, a putative ribosome binding site (RBS), the oligonucleotides used for sequencing and the locations of restriction sites referred to in the text are shown in Figure 4.5.

The sequencing reactions and related gels were performed and run as detailed in Materials and Methods (Section 2.3.17). All of the templates were sequenced an average of three times. The autoradiographs were read by eye and entered into computer files using the sequence assembler/editor program SEQED (Devereux *et al.*, 1984). Each autoradiograph was interpreted at least twice, each reading being entered into independent files. Duplicate files were then compared using the GAP program to eliminate "input" errors.



*BclI*

1 TGATCACGGTAGCAAGGGCGGCCACGCAGGGTCAGGGCTTGACGGCGACTATCAGCCAGT 60  
 ACTAGTGCCATCGTTCCCGCCGGTGGTCCCGAGTCCCGAACTGCCGCTGATAGTCGGTCA  
 EndProLysValAlaValIleLeuTrpAsp -

1 CACCGAGGAGGTTGTGATGCGCTCCGCCTGCCGCACCCGGAAGCCGGCCTCCTCGATCT 120  
 GTGGCTCCTCCAACAGCTACGCGAGGCGGACGGCGTGGGCCTTCGGCCGGAGGAGCTAGA  
 GlyLeuLeuAsnAspIleArgGluAlaGlnArgValArgPheGlyAlaGluGluIleGlu -

1 CCGCGCGGTACTCCTCGACGCTGCTGGACGAGCTGTGCGCCGCGGATGAGGCCGTGGTTGA 180  
 GCGCGGCCATGAGGAGCTGCGACGACCTGCTCGACAGCGGCGCCTACTCCGGCACCAACT  
 AlaArgTyrGluGluValSerSerSerSerSerAspGlyArgIleLeuGlyHisAsnLeu -

*BamHI*

1 GCGGTTGCAGTACGTCGTGCACCGTCGTGGTGGCCGGATCCGTCATCAGGTCGTAGACCA 240  
 CGGCAACGTCATGCAGCACGTGGCAGCACCCACCGGCCTAGGCAGTAGTCCAGCATCTGGT  
 ArgGlnLeuValAspHisValThrThrThrAlaProAspThrMetLeuAspTyrValVal -

1 CCAGCGCGCCGCGGAGGGCAGCGCCTCGTGGATGCGCGCGAGGAGCTTGCGCCGCTGCG 300  
 GGTGCGCGCGCGGCCCTCCCGTCCGCGGAGCACCTACGCGCGCTCCTCGAACGCGGCGACGC  
 LeuAlaGlyGlySerProLeuAlaGluHisIleArgAlaLeuLeuLysArgArgGlnPro -

*KpnI*

1 GTACCGGCCAGTCCGGCAGTACGTGCCCGGTGACCACGGCGTCCGCGGACGGCAGCGGGT 360  
 CATGGCCGGTCAGCCCGTCATGCACGGGCCACTGGTGCCGCGAGCGCCTGCCGTGCCCCA  
 ValProTrpAspProLeuValHisGlyThrValValAlaAspAlaSerProLeuProAsp -

1 CCGTCAGGAAGTCGCCGCCGTGGAAGCGGACGCGGCCTTCGAGGCCGCGCTCACGGATCA 420  
 GCCAGTCCTTCAGCGGCGGCACCTTCGCCTGCGCCGGAAGCTCCGGCGCGAGTGCCTAGT  
 ThrLeuPheAspGlyGlyHisPheArgValArgGlyGluLeuGlyArgGluArgIleLeu -

421 GCTCGCCGCCAGCGGGGCCAGCGCGGGCAGGTGGAACACGGCGCCGTGCAGGTGCGGGT 430  
 CGAGCGGCGCGGTGCGCCCGGTGCGCGCCGTCCAGCTTGTGCCGCGGCACGTCCACGCCCA  
 GluGlyAlaLeuProAlaLeuAlaProLeuAspPheValAlaGlyHisLeuHisProHis -

431 GGGCCAGGGCCACCCGCGTGGCGAGGTTGCCCCGTGCGCCGCCGATGTCGATGAAGGAGC 540  
 CCCGGTCCCGGTGGGCGCACCGCTCCAACGGGGCACGCGGCGGCTACAGCTACTTCCTGG  
 AlaLeuAlaValArgThrAlaLeuAsnGlyArgAlaGlyGlyIleAspIlePheSerGly -

## OLIGO. 907a

541 CGTACCCGCTCCAGTCCAGCGCGCGGCCAGCTCCTCGGCCGTGAAGGAGCTGAAGGTGT 600  
 GCATGGCGGAGGTCAGCTGCGCGCGCGGTGGAGGAGCGGCACTTCCTCGACTTCCACA  
 TyrGlySerTrpAspValArgAlaLeuGluGluAlaThrPheSerSerPheThrAsp -

## OLIGO. 807b

601 CGAAGTGCGCCATGACCTGACGGGCGCGCTCCGGGTCTCGTAGTGCTTGGGGTACGCCT 660  
 GCTTCACGCGGTACTGGACTGCCCGGGCGAGGCCAGGAGCATCACGAACCCCATGCGGA  
 PheHisAlaMetValGlnArgAlaArgGluProAspGluTyrHisLysProTyrAlaGlu -

661 CGGGCCCGTGGCTTCGGCGCCGGAACCCGCCTCGCCCTCCTGGAGCGCCGTCGTCAGCC 720  
 GCCCGGGCACGCGAAGCCGCGGCCTTGGGCGGAGCGGGAGGACCTCGCGGCAGCAGTCGG  
 ProGlyHisAlaGluAlaGlySerGlyAlaGluGlyGluGlnLeuAlaThrThrLeuGly -

## OLIGO. 907b

721 CGGCCCATACGTGGTATCTGCTTGGCGCGCTGCTGGAAGGACGGTGCCGCCCAGGTATTCCG 780  
 GCCGGGTATGCACCATCACGAACGCCCGCACGACCTCCTGCCACGGCGGGTCCATAAGCC  
 AlaTrpValHisTyrHisLysArgAlaHisGlnLeuValThrGlyGlyLeuTyrGluSer -

781 AGGCGCCGGACACCAGGAACGCCTGGGCGGCGGGGGAGTTGCGGTAGCGGTGCGCGTCCG 840  
 TCCGCGGCCTGTGGTCCTTGCGGACCGCGCTGCGGCTCAGCCCATCGCCAGCGGCAGGC  
 AlaGlySerValLeuPheAlaGlnAlaAlaProSerAsnArgTyrArgAspGlyAspAla -

## OLIGO. 807a

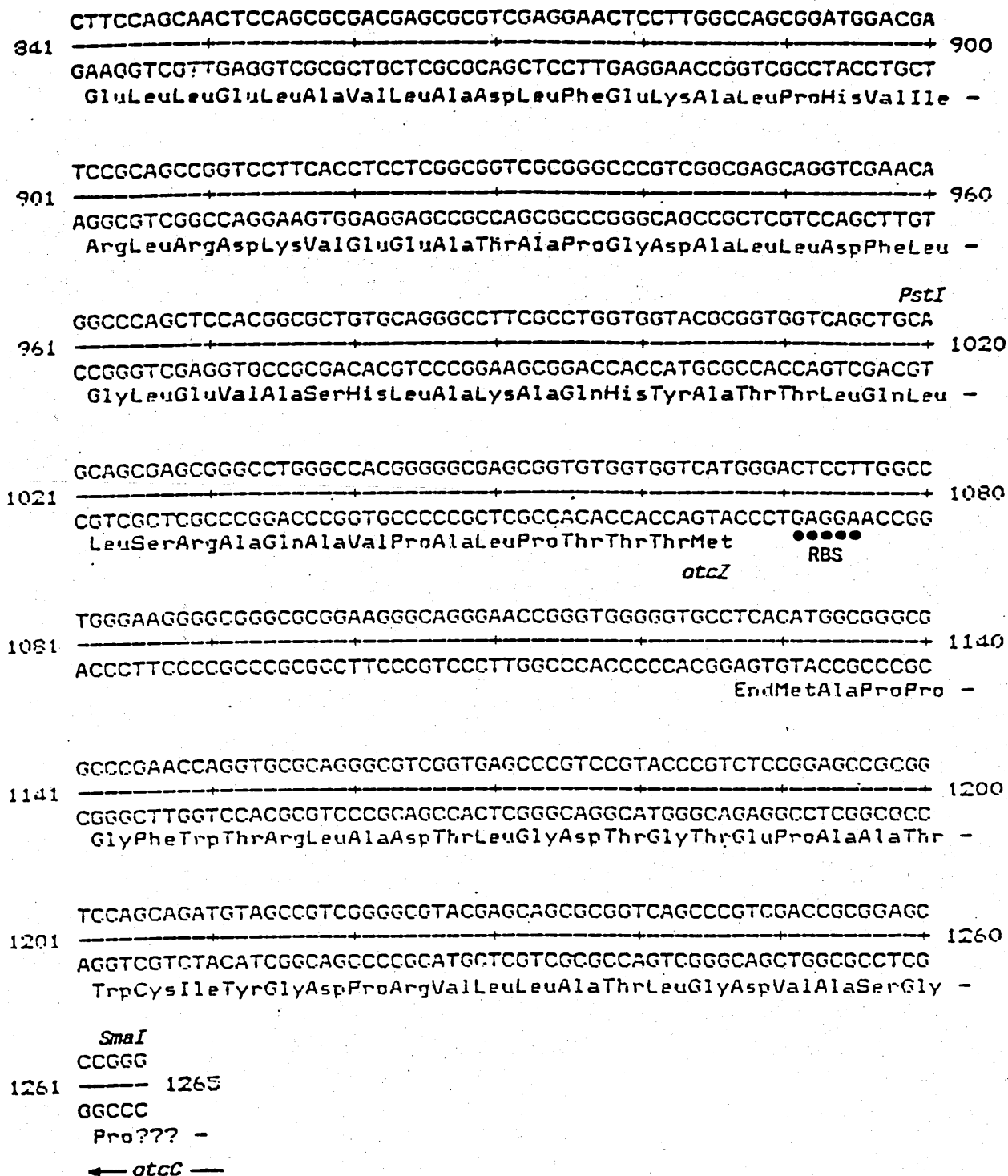


Figure 4.5. Annotated sequence of the 1,265bp *Bcl*I-*Sma*I region. The predicted amino acid sequences of *otcZ* and the C-terminus of *otcC* are shown below the corresponding nucleotide sequences. The shaded boxes ( ) indicate the nucleotides incorporated into the oligonucleotides which were used as primers for sequencing. A possible ribosome binding site (RBS) for *otcZ* is also indicated. For clarity, only restriction sites referred to in the text are indicated.

### 4.2.3 Computer-Assisted Sequence Analysis

Computers have an integral role in the analysis of DNA sequence, due to the enormous amount of data that has been generated and the complex algorithms employed in sequence analysis and comparison. The 1.3kb *BcII-SmaI* sequence was analysed using the "Sequence Analysis Software Package, Version 6.2" from the Genetics Computer Group at the University of Wisconsin (UWGCG) (Devereux *et al.*, 1984). This package has been compiled specifically for molecular geneticists wishing to analyse and compare sequence data rapidly and accurately. Access to the three main databases; EMBL, NBRF and GenBank was available through this system.

The UWGCG package was used successfully to predict the potential protein coding region of *otcZ* and the amino acid sequence of the gene product, to identify similarity with other proteins and consequently to suggest a possible enzymatic function for the predicted protein. The individual UWGCG programs which were used are listed below.

**WORDSEARCH:** searches for any sequence similarity between the query sequence and entries in the databases using a Wilbur and Lipman algorithm (1983).

**SEGMENTS:** aligns and displays segments of similarity found by WORDSEARCH.

**TFASTA:** translates nucleotide sequences held in a database in all six frames, before searching for any sequence similarity to a query peptide sequence using a Pearson and Lipman algorithm (1985).

**COMPARE:** compares two protein sequences for points of similarity.

**DOTPLOT:** displays graphically the points of similarity found by COMPARE.

**FOLD:** finds an optimum secondary structure for RNA molecules using the method of Zucker (1981).

**STEMLOOP:** finds inverted repeats within a sequence which could form secondary structure.

Before the analysis of the *otcZ* sequence is described, it is appropriate that a few of the terms used commonly in the proceeding sections are explained. An "algorithm" is defined as the description of how to solve a problem presented in terms that could be solved using a computer. DNA sequence, in common with its biological role, is analysed as a linear array of information. A "window" is a fixed length of sequence within which a statistical calculation is performed. This calculation is specified by the algorithm. A "scoring matrix" is a two dimensional array of scores which represent the degree of "relatedness" between entities being compared. For example, the Dayhoff matrix, which is used by COMPARE, contains values which represent the degree of "evolutionary relatedness" for every possible pairwise comparison of amino acids. Matrices are extremely useful for identifying imperfect homologies between amino acid sequences.

#### 4.2.3.1 Potential Protein Coding Regions

An open reading frame (ORF) is defined as a string of triplets, which does not contain any translation stop codons (i.e UAA, UGA or UAG). Only ORFs which include an in-frame translation start codon at the 5' end can be considered as potential protein coding regions (PPCRs). In *Streptomyces*, as in *E. coli*, the most common initiation codon is AUG (82%). Not surprisingly perhaps, the GUG codon is noticeably more frequent in *Streptomyces*, being present in approximately 18% of genes compared to the 3% in *E. coli*. Translation initiation at other codons has not been identified in *Streptomyces* genes. In *E. coli* a very small number of proteins are initiated, however, by the UUG codon and the AUU codon. (Gold *et al.*, 1981; Stromo, 1986). Of the *Streptomyces* genes which have been sequenced completely, 67% are terminated by UGA, 29% by UAG and 4% by UAA (Bibb, pers. comm.).

Potential protein coding regions (PPCRs) in the 1.3kb *BcII*5-*Sma*I9 fragment were identified using the FRAME algorithm (Bibb *et al.*, 1984). FRAME analysis exploits the non-random distribution of bases within codons to identify PPCRs. *Streptomyces* DNA has typically a base composition of 73% mol G+C (Enquist and Bradley, 1971), which is reflected in the biased usage of synonymous codons. The average mol G+C composition in the first, second and third positions of streptomycete codons are 70%, 45% and 90%, respectively (calculated from a file of 63

genes compiled by Bibb, unpublished result). Intervening regions between coding sequences have a random distribution of bases within consecutive triplets. The FRAME program scans DNA sequence and determines the mol% G+C composition at the first, second and third positions of a number of triplets from a "window", which usually contains 25 to 50 triplets. The mol% G+C for each of the three base positions within the "window" are then plotted against a linear representation of the sequence, above the first nucleotide of the "window". The "window" is then moved along one triplet and the calculation repeated. This procedure is continued until the entire sequence has been analysed.

FRAME revealed two PPCRs within the 1.3kb *Bcl*I-*Sma*I region (Figure 4.6). The heavy line is the plot of the mol% G+C composition at every third base, in a window of 50 triplets, starting at base 3 of the *Bcl*I-*Sma*I sequence. Similarly, the beaded (-●-) line is the plot starting at base 2 and the thin line is the plot starting at base 1. The analysis of the graphic output will be considered from the *Bcl*I site (base 1). The FRAME plots are immediately diverse and start to converge around position 920. The heavy line plot commencing from base 3, averages around 45% mol G+C and thus, corresponds to the second position of the codons in the PPCR. Similarly, the beaded line originating from base 2, with its average G+C composition of 95%, corresponds to the third codon position and the thin line from base 1, which fluctuates around 75%, corresponds to codon position 1. Therefore, the open reading frame for PPCR1 is translated from right to left in the *Bcl*I-*Sma*I sequence in nucleotide frame 2 (N2<).

Two potential start codons, a AUG at 1066nt and a GUG at 1042nt and a UGA stop codon at 34nt respectively, were identified for PPCR1 (graphically represented above the FRAME plot). The AUG codon is preceded by a reasonable ribosome binding site (AAGGAG) with a high degree of complementarity to a region close to the 3' end of 16S rRNA from *S. lividans* (sequenced by Bibb and Cohen, 1982) (Figure 4.5). This putative RBS is separated from the AUG triplet by 4nt and is well within the limits of 5-9nt for Shine-Dalgarno interactions (Gold *et al.*, 1987). No significant complementarity to the 3' terminus of *S. lividans* 16S rRNA was found within the sequence immediately upstream of the GUG

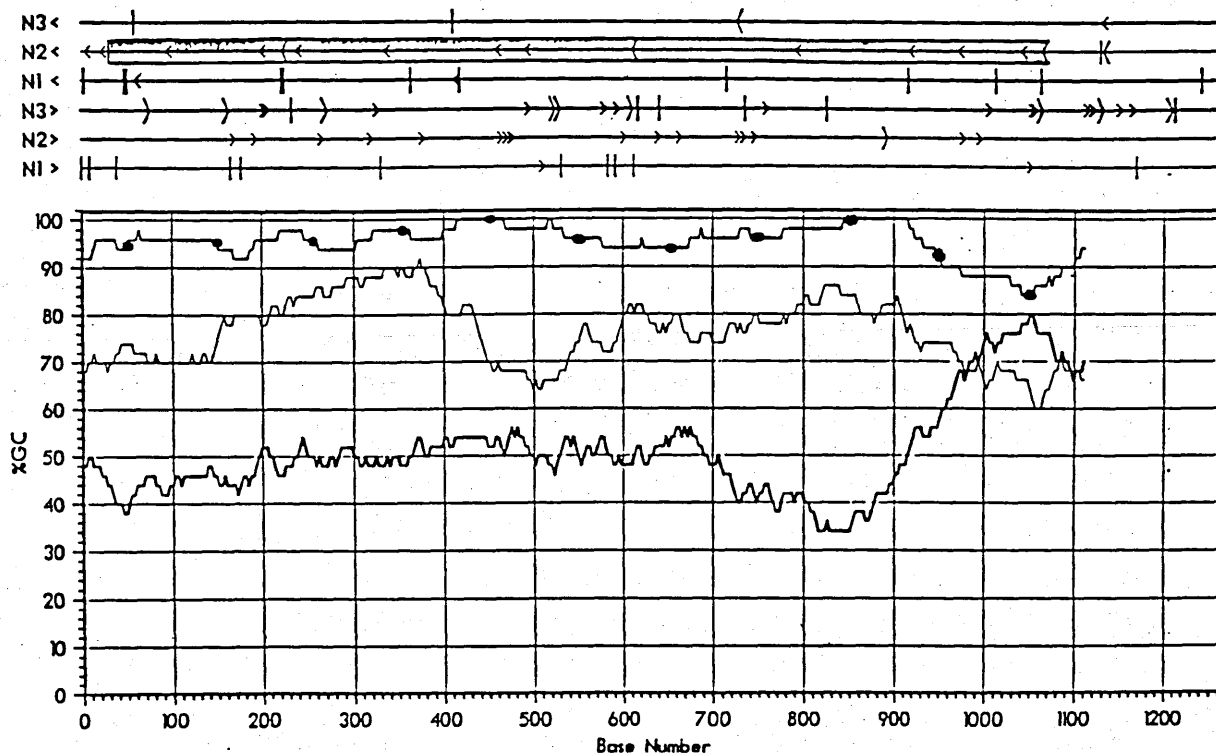


Figure 4.6. FRAME plot (Bibb *et al.*, 1984) of the 1,265bp *otcZ* region using a "window" of 50 triplets. The thin line plot (—) represents the first triplet position and its associated reading frames (N1> and N2<); the beaded line (—●—) represents the second triplet position and its related reading frames (N2> and N3<) and the heavy line (—) indicates the third triplet position and its associated reading frames (N3> and N1<). Symbols > and < indicate AUG codons in the forward and reverse frames respectively, Symbols > and < indicate GUG codons and the symbol | indicates stop codons.

codon. It was concluded, therefore, that the AUG codon at 1066nt was the probable translation initiation codon for PPCR1. Translation initiation is considered in more detail within Chapter 7. The FRAME analysis starting at the *Bcl*I site was immediately diverse, as the non-coding region between the *Bcl*I site and the 3' end of *otcZ* was very short (31bp) and consequently the "windows" (150nt) that covered this region were primarily influenced by the coding sequences of PPCR1.

Although the plots started to converge at base position 920, an equilibrium position was never reached around 73 mol% G+C composition. Instead the plots diverged once more. This divergence suggested the presence of a second short PPCR at the extreme right end of this region. Preliminary sequence data from the Pfizer Laboratories at Sandwich (Butler, unpublished result), has identified PPCR2 as the 3' end of the *otcC* gene, which encodes the ATC oxygenase (see Section 4.1). The translation stop codon for *otcC* is at base 1129, which is in the same frame as PPCR1. The intervening region of 60bp between PPCR1 and PPCR2 is much shorter than the window (150nt) used for FRAME analysis. Therefore, the FRAME values were strongly influenced by sequences in PPCR1 and PPCR2. This explains why an equilibrium position around 73 mol% G+C, which is normally typical of non-coding sequence, was never reached. The same PPCRs in the *Bcl*I-*Sma*I region were also identified using the TESTCODE program from the UWGCG package (Devereux, 1984).

The *otcZ*151 mutation, which was mapped close to the "left" of the *Pst*I site (Butler *et al.*, 1989), probably lies within PPCR1. Therefore, PPCR1 will be referred to as the *otcZ* structural gene. The *otcZ* gene was predicted to encode a protein of 344 amino acids with a molecular weight of 37.36kD. As expected for a streptomycete gene, *otcZ* has an average mol G+C distribution in the first, second and third codon positions of 75.6%, 46.4% and 94.5% respectively. Comparison of the codons used in *otcZ* with a codon preference table compiled from 63 *Streptomyces* genes (Bibb, unpublished data), revealed that the biased usage of synonymous codons in *otcZ* was in accordance with the codon preferences observed for other streptomycete genes (Figure 4.7).



(A)

	T	C	A	G	
T	F TTT 0.00	S TCT 0.00	Y TAT 0.00	C TGT 0.00	T
	F TTC 3.19	S TCC 2.61	Y TAC 2.61	C TGC 0.00	C
	L TTA 0.00	S TCA 0.00	* TAA 0.00	* TGA 0.29	A
	L TTG 0.29	S TCG 0.58	* TAG 0.00	W TGG 1.16	G
C	L CTT 0.00	P CCT 0.00	H CAT 0.29	R CGT 0.58	T
	L CTC 4.06	P CCC 2.61	H CAC 4.64	R CGC 4.06	C
	L CTA 0.00	P CCA 0.00	Q CAA 0.29	R CGA 0.00	A
	L CTG 8.41	P CCG 2.91	Q CAG 2.61	R CGG 2.61	G
G	I ATT 0.00	T ACT 0.00	N AAT 0.00	S AGT 0.00	T
	I ATC 2.32	T ACC 2.90	N AAC 1.16	S AGC 1.74	C
	I ATA 0.29	T ACA 0.29	K AAA 0.00	R AGA 0.00	A
	M ATG 0.87	T ACG 2.03	K AAG 2.03	R AGG 0.00	G
A	V GTT 0.00	A GCT 0.00	D GAT 0.29	G GGT 0.58	T
	V GTC 3.19	A GCC 7.54	D GAC 5.80	G GGC 6.09	C
	V GTA 1.16	A GCA 0.29	E GAA 1.16	G GGA 0.00	A
	V GTG 2.90	A GCG 5.80	E GAG 6.09	G GGG 1.74	G

(B)

	T	C	A	G	
T	F TTT 0.04	S TCT 0.06	Y TAT 0.11	C TGT 0.10	T
	F TTC 2.89	S TCC 2.11	Y TAC 2.22	C TGC 0.69	C
	L TTA 0.02	S TCA 0.15	* TAA 0.01	* TGA 0.26	A
	L TTG 0.26	S TCG 1.40	* TAG 0.05	W TGG 1.55	G
C	L CTT 0.22	P CCT 0.18	H CAT 0.16	R CGT 0.51	T
	L CTC 3.52	P CCC 2.36	H CAC 2.33	R CGC 3.74	C
	L CTA 0.03	P CCA 0.07	Q CAA 0.18	R CGA 0.32	A
	L CTG 5.08	P CCG 2.87	Q CAG 2.45	R CGG 2.96	G
G	I ATT 0.14	T ACT 0.14	N AAT 0.10	S AGT 0.14	T
	I ATC 3.07	T ACC 4.49	N AAC 2.35	S AGC 1.46	C
	I ATA 0.10	T ACA 0.15	K AAA 0.12	R AGA 0.10	A
	M ATG 1.55	T ACG 1.99	K AAG 2.29	R AGG 0.46	G
A	V GTT 0.27	A GCT 0.38	D GAT 0.27	G GGT 0.74	T
	V GTC 4.40	A GCC 7.77	D GAC 6.16	G GGC 0.74	C
	V GTA 0.16	A GCA 0.55	E GAA 1.11	G GGA 0.85	A
	V GTG 3.08	A GCG 4.39	E GAG 4.83	G GGG 1.70	G

Figure 4.7. (A) Codon usage in the *otcZ* structural gene (B) Codon usage in a file of 63 streptomycete genes (compiled by Bibb, unpublished data). The numbers represent the frequency with which the codons are used. The frequency is expressed as a percentage (%) of the total number of codons in the file.

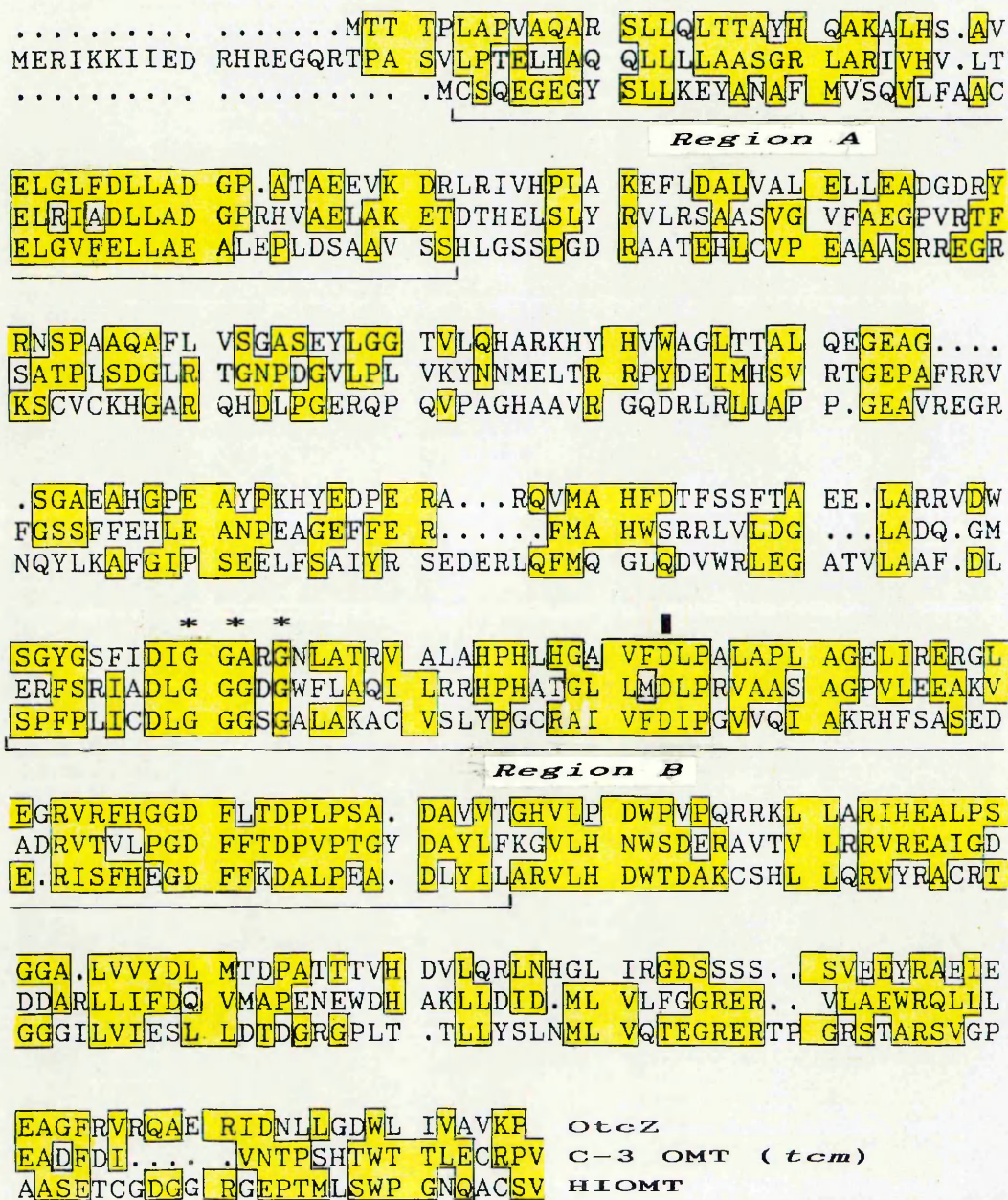
#### 4.2.3.2 Amino Acid Sequence Comparisons

Using WORDSEARCH and SEGMENTS, comparison of the predicted amino acid sequence of *otcZ* to protein sequences held in the NBRF-Protein database (Release 24, March 1990), revealed that a hydroxyindole O-methyltransferase (HIOMT) from bovine pineal glands (Ishida *et al.*, 1987) had the highest sequence similarity. Comparison of the amino acid sequence of *otcZ* with entries in the EMBL (Release 19, July 1990) and GenBank (Release 60, June 1989) databases using TFASTA did not identify any other entries with sequence similarity to the *otcZ* gene product (OtcZ). The bovine hydroxyindole O-methyltransferase is a protein of 350 amino acids with a molecular weight of 37.78kD, similar to that of OtcZ. HIOMT is involved in the biosynthesis of melatonin, which has been studied intensively by biochemists and physiologists. The amino acid sequence of HIOMT was predicted from the DNA sequence of a cDNA clone isolated by screening a library, which had been constructed in the lambda gt expression vector, with monoclonal antibodies to purified HIOMT (Ishida *et al.*, 1987).

The *tcmIa* ORF4, from the tetracenomycin C-producer, *S. glaucescens*, was reported by Hutchinson at the 19th UCLA Symposium in Colorado (1990) to encode a protein which also has sequence similarity to the bovine HIOMT. The predicted product of *tcmIa* ORF4 has a molecular weight of 55.93kD and contains 494 amino acids (Hutchinson, pers. comm.). The sequence similarity is limited to the carboxy-terminal two-thirds of HIOMT. However, the amino-terminal region, extending to one-third of the predicted product of *tcmIa* ORF4, has high sequence similarity to the amino-terminal half of the predicted products of *act* ORF4 and *gra* ORF4 and the full-length product of *whiE* ORF6 (Hopwood and Sherman, 1990). Both *act* ORF4 and *gra* ORF4 are postulated to encode bifunctional cyclase/dehydrases (Hopwood and Sherman, 1990), which are required for the formation of the correct isochromanquinone nucleus, as they complement *actVII* mutants that accumulate an aberrantly-cyclised shunt product, mutactin (Zhang *et al.*, 1990). The amino-terminus of the predicted protein product of *tcmIa* ORF4 and the product of *whiE* ORF6 may also contain cyclase activity. The reader is directed to the Introduction (Section 1.12) for a more detailed discussion of the proposed role of the above gene products in the synthesis of polyketides.

Production of tetracenomycin C was restored when *tcmIa* ORF4 was introduced on a plasmid into *tcmII* and *tcmV* mutants, which accumulate C-3 demethyl intermediates (Hutchinson, pers. comm.) The gene product of *tcmIa* ORF4 appeared, therefore, to have C-3 O-methyltransferase activity, in addition to its putative role in cyclization. Recently, Hutchinson and coworkers were surprised to find that deletion of the proposed ribosome binding site and translation initiation codon for *tcmIa* ORF4, did not interfere with the complementation of *tcmII* and *tcmV* mutants (Hutchinson, pers. comm.). An AUG codon, which is located approximately one-third of the way into *tcmIa* ORF4 (amino acid 137), is preceded by a reasonable ribosome binding site and may act as an initiation codon for the translation of the C-3 O-methyltransferase. The boundary between the putative "cyclase" and "methyltransferase" coding regions is currently being re-sequenced and expression studies carried out to establish if *tcmIa* ORF4 is, in fact, two juxtaposed PPCRs. For the purposes of comparison with HIOMT and OtcZ, only the predicted product of *tcmIa* ORF4 from methionine 137 (OMT) was considered. Optimum alignment of the amino acid sequences of OtcZ, HIOMT and *tcmIa* OMT by rationalising all the possible pairwise comparisons using WORDSEARCH and SEGMENTS (Figure 4.8), revealed that OtcZ has 20.3% (71/350) and 23.5% (84/358) overall amino acid sequence identity to HIOMT and *tcmIa* OMT respectively.

COMPARE and DOTPLOT were used to identify the most conserved regions in the above protein sequences. COMPARE analyses two protein sequences, creating a file of points of similarity between them which can be displayed using DOTPLOT. The COMPARE program finds points of similarity using a "window"/"stringency" match criteria. The sequences are compared in all possible registers and where the quality of the matches within a "window" of amino acids is greater than or equal to a value specified by the "stringency", the point is written to a file. Matches are determined using the Dayhoff matrix which contains a numerical value representing the evolutionary distance for every possible amino acid comparison (Devereux *et al.*, 1984).



**Figure 4.8.** Alignment of the predicted amino acid sequences of *tcmla* OMT, *OtcZ* and *HIOMT*. Lines (|---|) indicate the two regions identified by COMPARE and DOTPLOT as being the most evolutionarily conserved (Figure 4.9). Asterisks (\*) and the symbol (■) indicate the positions of the conserved glycines and aspartate, respectively, which are thought to be involved in the binding of S-adenosylmethionine. Amino acids were grouped as follows; (GASTP), (ILVM), (FYW), (QNED), (HKR) and (C).

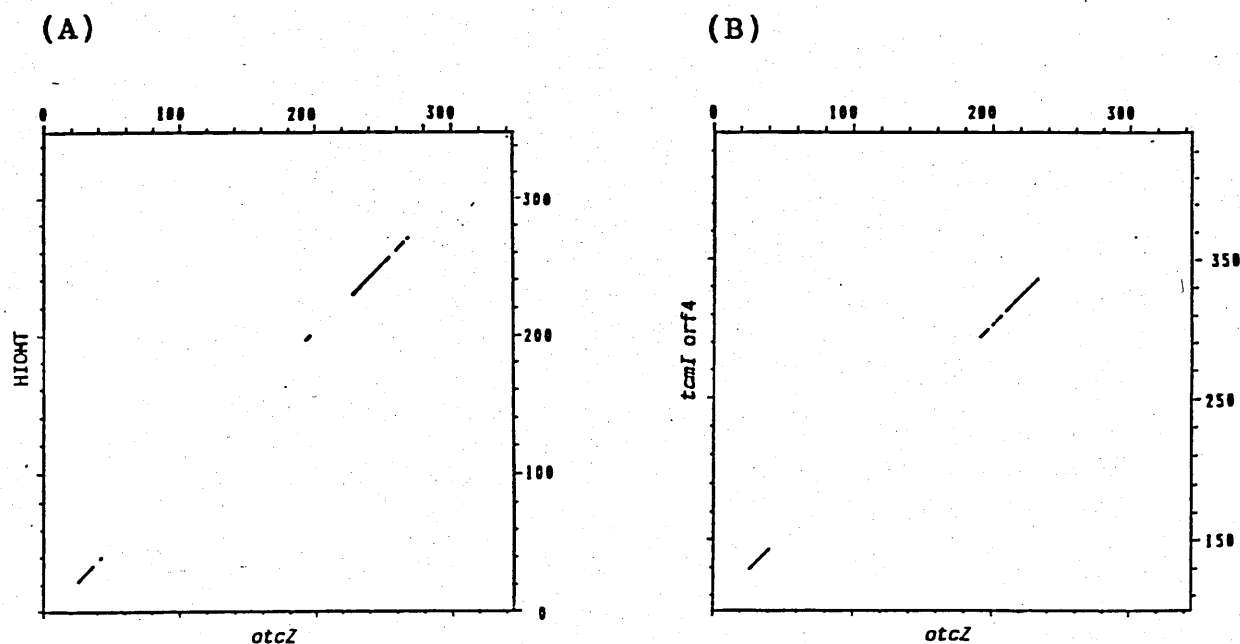


Figure 4.9. COMPARE/DOTPLOT analyses of the deduced protein product of *otcZ* versus (A) HIOMT (Ishida *et al.*, 1987) and (B) the deduced protein product of *tcmIa* ORF4 (amino acids 137-494) (Hutchinson pers. comm.). The stringency was 20 for a window of 30 amino acids (for definitions of window and stringency, see Devereux *et al.*, 1984).

Using these match criteria it is possible to determine imperfect homologies and show evolutionary relationships. COMPARE puts only one point in the file at the middle position of the window whenever the stringency is met.

DOTPLOTS of the comparisons between OtcZ, HIOMT and *tcmIa* OMT highlighted two regions, shared by all three proteins, which appeared to be particularly well conserved. Within region A, OtcZ (amino acids 6 to 53) shows 22.0% (11/50) and 36.7% (18/49) amino acid sequence identity to HIOMT (2-51) and *tcmIa* OMT (159-207) respectively. Within Region B, OtcZ (amino acids 171 to 246) has 34.7% (26/75) and 37.0% (27/73) amino acid sequence identity to HIOMT (175-247) and *tcmIa* OMT (324-398) respectively (Figure 4.8).

S-Adenosylmethionine is the mononucleotide coenzyme that donates the methyl group in the hydroxyindole O-methyltransferase reaction (Axelrod and Weissbach, 1961; Nakane *et al.*, 1983)). A "fingerprint" sequence has been identified that is characteristic of the nucleotide-binding domains of many proteins (Wierenga and Hol, 1983). Although the nucleotide-binding sequence was identified initially in proteins that bind NAD<sup>+</sup> (Wierenga and Hol, 1983), this motif has also been found in proteins which bind nucleotides such as FAD, GTP and ATP (Ogawa *et al.*, 1987).

X-ray crystallographic studies, have shown that amino acids within the "fingerprint" sequence interact with the adenosine 5' diphosphate moiety in the nucleotides (Wierenga and Hol, 1983). The dinucleotide-binding domain consists of a region of  $\beta$ -sheet followed by a region of  $\alpha$ -helix and then another region of  $\beta$ -sheet, making up a  $\beta\alpha\beta$  unit with the following features: (i) the sequence G-X-G-X-X-G in the region joining the first  $\beta$ -strand and the amino terminus of  $\alpha$ -helix, (ii) the presence of a hydrophilic residue at the beginning of the first  $\beta$ -sheet region and (iii) the presence of an acidic residue at the carboxyl terminus of the second  $\beta$ -sheet. In addition, several hydrophobic residues are located at specific positions in the hydrophobic core of the  $\beta\alpha\beta$  unit. The first and third invariant glycines

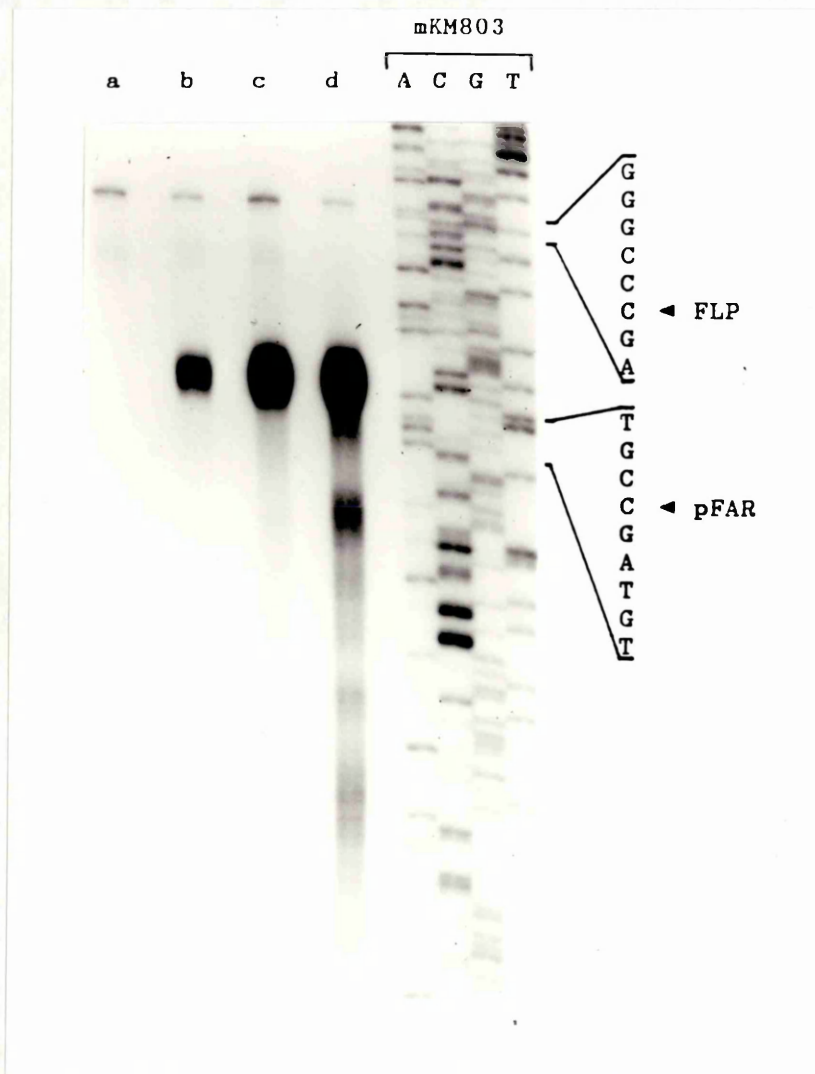
of the nucleotide-binding  $\beta\alpha\beta$  units, allow the polypeptide chain to make a sharp turn between the first  $\beta$ -strand and the  $\alpha$ -helix. The presence of a side chain at the first glycine position would interfere with the binding of the ribose moiety. The second glycine allows the pyrophosphate moiety to approach the amino terminus of the  $\alpha$ -helix, which it binds through hydrogen bonding. For an extensive comparison of nucleotide-binding domains from proteins with diverse functions, the reader is directed to Ogawa *et al.* (1987).

Examination of the alignment of the amino acid sequences of OtcZ, *tcmIa* OMT and HIOMT, has identified a "consensus" sequence within Region B that resembles the nucleotide-binding "fingerprint" and may form a S-Adenosylmethionine (SAM)-binding domain (Figure 4.8). Owing to the speculative nature of the UWGCG programs available for predicting secondary structure, it was not possible to establish if this "consensus" sequence forms a  $\beta\alpha\beta$  domain. The putative SAM-binding sequence contains three invariant glycines (G-G-X-X-G), which are present in the "fingerprint" sequence for NAD<sup>+</sup>-binding proteins (Wierenga and Hol, 1983). The absence of an amino acid, separating the first and second glycines of the putative SAM-binding domain, may reflect the absence of a pyrophosphate moiety in S-Adenosylmethionine. A conserved aspartate residue is located 19 residues towards the carboxyl-terminus from glycine 3 (Figure 4.8). In the corresponding position of the "fingerprint" sequence, an acidic residue, such as aspartate, is proposed to form strong hydrogen bonds with the 2'-hydroxyl group of the adenine-ribose. It should be noted that the assignment of a SAM-binding sequence to region B is highly speculative, as SAM-binding domains remain to be characterised by X-ray crystallography and molecular modelling. To date, the analysis of amino acid sequence in region A has not identified any similarity to consensus motifs for groups of functionally-related proteins.

#### 4.2.4 Analysis of the Transcription of *otcZ*

To establish if transcription originated from a promoter located close to the predicted translation initiation codon of *otcZ*, S1 nuclease protection analysis was undertaken (Figure 4.10) using a single-stranded





**Figure 4.10.** High resolution S1 mapping of transcripts within the *Pst*I-*Sma*I. Single-stranded probe ( $1.2 \times 10^7$  Cerenkov cpm.pmol $^{-1}$ ) was generated from mKM803 by extension from the universal primer (-20) and digested with *Eco*R1 (see Materials and Methods 2.3.19). The probe (0.1pmoles) was hybridized at 52°C with ca. 10µg of total RNA from the following (a) *S. rimosus* M15883S, (b) *S. rimosus* M4018 grown on Tryptone Soya Broth (16hrs), (c) *S. rimosus* M4018 producing OTC (ca. 60µgml $^{-1}$ ) on Liquid Complete Medium (56hrs) (see Figure 3.4) and (d) *S. rimosus* M15883 producing OTC (ca. 0.7mgml $^{-1}$ ) on Trusoya Medium 1 (46hrs) (see Figure 3.6). The sequencing ladder was produced from mKM803 (see Section 4.2.2) by extension from the -40 primer. The position in the mKM803 sequencing ladder that is coincident with the RNA-protected fragment derived from transcripts originating far upstream of the *Sma*I site is indicated by pFAR. The position of probe protected along its entire length by contaminating template DNA is also indicated (FLP).



probe that included 250bp of sequence from the *Pst*I<sub>8</sub> to the *Sma*I<sub>9</sub> site. The predicted start codon for *otcZ* is 51bp upstream of the *Pst*I<sub>8</sub> site. The probe was generated from mKM803 template, which was used previously for sequencing (Figure 4.3), by extension from the universal primer (-20) and digestion with *Eco*RI (Materials and Methods 2.3.18). Nucleotides derived from M13mp18 sequence accounted for 32nt at the 5' end and 14nt at the 3' end of the 297nt probe respectively.

The probe was hybridized with a selection of RNA samples isolated from *S. rimosus* in which the relative abundances of transcripts from *otrA*p1 and pFAR had been shown to be very different (Section 3.2.4). Transcripts originating from *otrA*p1 were by far the most abundant (>95%) in total RNA isolated from M4018 grown on Tryptone Soya Broth, while transcripts from *otrA*p1 and pFAR had similar abundances in RNA isolated from M4018 grown in Liquid Complete Medium for 56hrs. In total RNA isolated from M15883 grown on Trusoya Medium 1 for 46hrs, only transcripts originating from pFAR could be detected. The probe was also hybridized with total RNA isolated from *S. rimosus* M15883S, as a negative control.

A single fragment was protected by transcripts in all of the above RNA samples (except the negative control), which co-migrated with position 200 (numbering from the *Pst*I site) in the sequencing ladder. As mKM803 was sequenced using the -40 primer, the extension products, which formed the sequencing ladder, contained 52 nucleotides of M13mp18 sequence at the 5' end. Therefore, the RNA-protected fragments were 252nt long. This length of protection corresponds almost exactly to that predicted for transcripts extending through the entire 250bp *Pst*I<sub>8</sub>-*Sma*I<sub>9</sub> region. The absence of a RNA-protected fragment in the M13883S sample indicated that the transcript(s) providing protection were specific to the oxytetracycline cluster.

In all four samples, including the control, a fragment of 297 nucleotides was detected that co-migrated with position 245 in the sequencing ladder. This length of protection corresponds to full-length probe. Protection of the probe over its entire length probably results from contamination of the probe with a small amount of complementary template DNA.

### 4.3 Discussion

Methylation is believed to occur at two steps in oxytetracycline biosynthesis; (i) the C-methylation of the virgin tetracyclic nucleus to form 6-methylpretetramid and (ii) the N-methylation of 4-amino anhydrotetracycline (4-amino ATC) to form anhydrotetracycline (ATC) (Figure 4.11) (Rhodes *et al.*, 1981). The latter methylation is blocked in the *otcD4* mutant, which was able to convert purified ATC, but not 4-amino ATC, to oxytetracycline (OTC). The *otcZ151* mutant could convert 4-amino ATC to OTC and thus, could not be blocked in N-methylation (Rhodes *et al.*, 1981).

It is predicted that the *otcZ* mutant would accumulate 6-demethyl derivatives of oxytetracycline intermediates. Interestingly, purified 6-demethylchlorotetracycline (6-demethyl CTC) and 6-demethyltetracycline (6-demethyl TC) from a mutant of *S. aureofaciens* were shown to have similar levels of *in vivo* antimicrobial activity as chlorotetracycline and tetracycline respectively (McCormick *et al.*, 1957). This suggests that the absence of a methyl group on C-6 is not sufficient, on its own, to explain the antibiotic-negative phenotype of the *otcZ151* mutant. However, the detailed chemical analysis of mutants of *S. glaucescens* blocked in tetracenomycin C production has shown that mutations which block the addition or modification of a particular side group can have "knock on" effects at later stages in the biosynthesis of this antibiotic (Motamedi *et al.*, 1986; Yue *et al.*, 1986). For example, *tcmII* mutants, which are thought to be deficient in C-3 O-methyltransferase activity (Section 4.2.3.2), accumulate intermediates and shunt products that are not oxidised at C-2, C-4a and C-12a, as well as not being methylated at C-3. In the article by McCormick (1975), a full description of the mutant of *S. aureofaciens* was not provided. It is possible that a range of 6-demethyl intermediates were produced. The C-6 methyl group may be important for substrate recognition by enzymes which act later in the pathway and modify or add groups required for antimicrobial activity. It is also possible that the absence of a 6-methyl group could affect the reactivity of some of the later steps (Stark, pers. comm.). Analysis of the ATC oxygenase from *S. rimosus* has revealed that the enzyme cannot hydroxylate 7-chloro-ATC, which is the natural substrate

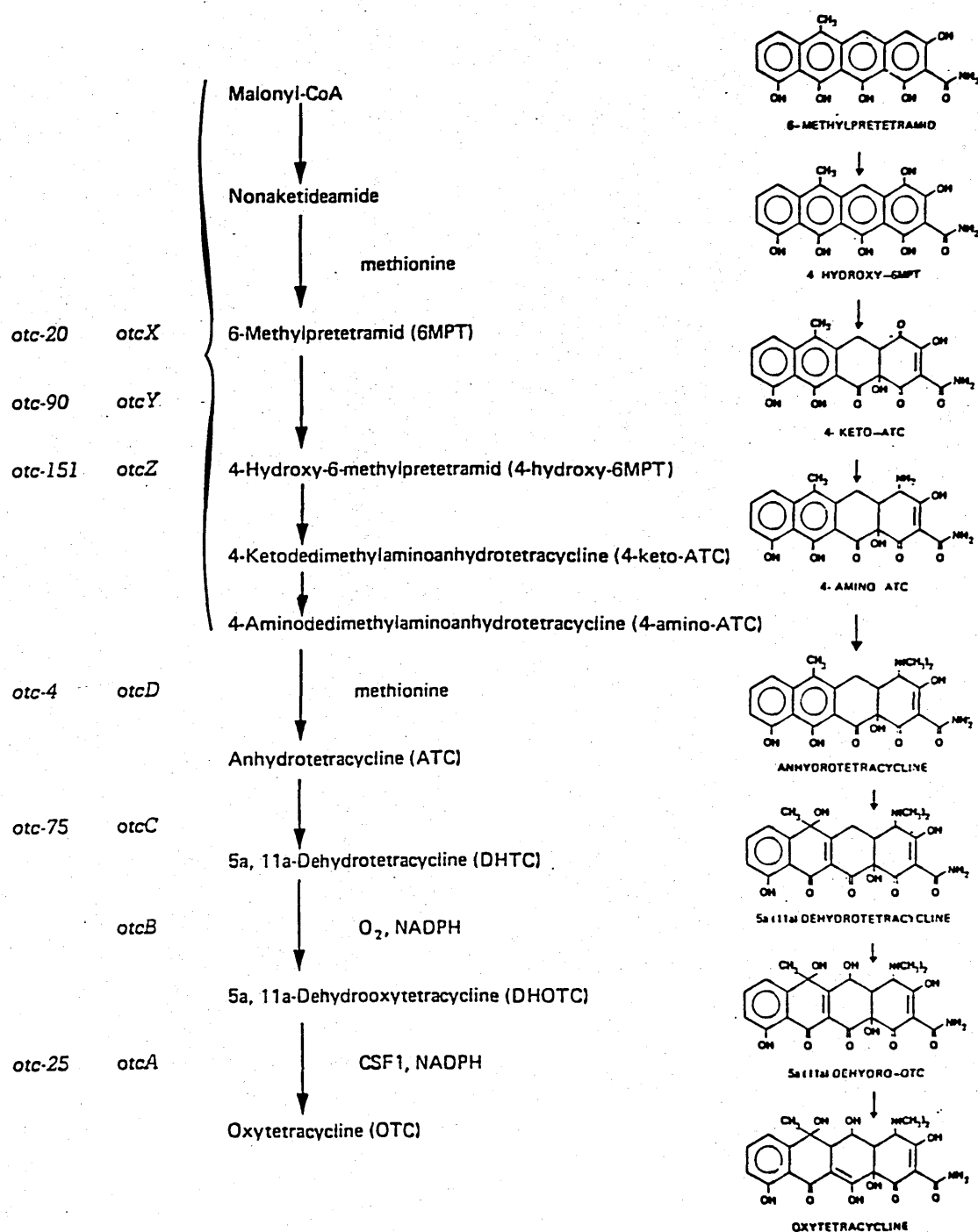


Figure 4.11. Biosynthesis of oxytetracycline (OTC). The steps at which various OTC-negative mutants are blocked are indicated. From Rhodes *et al.*, 1981.

of the enzyme from *S. aureofaciens* (Novotna, pers. comm.). It is possible that other subtle differences exist between the enzymology of tetracycline/chlorotetracycline production in *S. aureofaciens* and oxytetracycline biosynthesis in *S. rimosus*. Should the disruption of C-6 methylation in *S. aureofaciens* not hinder significantly the flux of intermediates through subsequent steps in CTC and TC biosynthesis, it is possible that the same situation might not be true of oxytetracycline production in *S. rimosus*. To date, there are no reports of mutants of *S. rimosus*, which produce 6-demethyloxytetracycline. Therefore, the proposal that the *otcZ* gene encodes a methyltransferase, which adds the C-6 methyl group to the tetracycline nucleus, is not inconsistent necessarily with the antibiotic-negative phenotype of the *otcZ151* mutant. Chemical analysis of intermediates and shunt products accumulated in antibiotic-negative mutants of *S. rimosus* would complement the detailed genetic analysis of the oxytetracycline cluster.

In the previous chapter, transcriptional analysis indicated that a promoter beyond the *Pst*I site was responsible for the transcription of the *otrA* resistance gene during the production of oxytetracycline. The limited transcriptional analysis described in this chapter suggests that there is no promoter within the *Pst*I-*Sma*I region, between the end of *otcC* and the start of *otcZ*. It would appear, therefore, that *otrA*, *otcZ* and possibly *otcC* are transcribed as part of a polycistronic message during antibiotic production. The cotranscription of *otrA* with biosynthetic genes probably ensures that resistance to oxytetracycline increases in parallel with antibiotic biosynthesis. Furthermore, these results suggest that the delayed onset of oxytetracycline production can be attributed, at least in part, to the control of gene expression at the level of transcription. It remains a formal possibility, however, that the catalytic activity of some of the enzymes involved in the production of OTC may also be regulated. If the production of antibiotic is not synchronised, inducible resistance (discussed in Section 3.3) could be a means of protecting vegetative mycelium not devoted to antibiotic production from exogenous tetracycline produced by mycelium in other regions of the same colony or other colonies. The transcription of the "left" end of the oxytetracycline cluster is considered further in the following chapter.

**CHAPTER 5**  
**CHARACTERISATION OF**  
**THE *OTCC/OTCX* PROMOTER REGION**

### 5.1 Introduction

The study of endospore formation in *B. subtilis* in response to nutrient deprivation has been a paradigm of developmental regulation in bacteria. Alternative sigma ( $\sigma$ ) factors and thus, alternative forms of RNA polymerase holoenzymes, which recognise different class of promoters, have an integral role in "re-programming" transcription during morphological development in this prokaryote. It has been implied often that alternative  $\sigma$  factors may have a similar role in the morphological and physiological development of *Streptomyces*. Recently, this hypothesis gained credence with the finding that the switch from the extension of the aerial hyphae to the development of chains of spores is controlled by a sigma factor ( $\sigma^{WhiG}$ ) (Chater *et al.*, 1989).

The expression of several antibiotic pathways would appear to be regulated by the controlled expression of ancillary factors, which either act along with the RNA polymerase to facilitate transcription at the end of the growth phase or antagonise RNA polymerase-promoter interactions during primary metabolism. Pathway-specific regulatory genes have been identified within the clusters of genes for the biosynthesis of actinorhodin (Rudd and Hopwood, 1979), methylenomycin (Chater and Bruton, cited in Hopwood *et al.*, 1986), bialaphos (Anzai *et al.*, 1987), streptomycin (Distler *et al.*, 1988), milbemycin (Malpartida *et al.*, 1987) and tylosin (Hershberger, 1989). These specific examples are discussed in more detail within the General Introduction, Section 1.6. At present, it is open to debate whether the "housekeeping" RNA polymerase or another form, associated with a minor  $\sigma$  factor, recognises the promoters of genes involved in antibiotic production.

It has been established for some time that the ratio of regulatory proteins to their cognate binding sites on DNA is of key importance in regulating gene expression (Ptashne, 1986). The molecular analysis of physiological and morphological differentiation in *S. coelicolor* A3(2) has provided several examples which illustrate this point. The introduction of extra copies of the *actII* region, which contains a gene encoding an activator of transcription, lead to the significant over-production of actinorhodin (Hopwood *et al.*, 1986). A fragment of DNA from *B. subtilis*

carrying a promoter recognised by  $\sigma^{Hic}$  caused reduced sporulation when introduced into *S. coelicolor* A3(2) on a high-copy-number plasmid (Chater *et al.*, 1989). The latter example is reminiscent of the reduction in antibiotic production, the "switch-off" phenomenon, which was observed when certain regions of the oxytetracycline cluster were introduced into *S. rimosus* M15883 on high-copy-number vectors (Butler *et al.*, 1989). The same regions did not "switch-off" oxytetracycline production when they were introduced into the same production strain on a low-copy-number vector. From these results, it could be interpreted that the regions which cause antibiotic production to be "switched off" contain a recognition site for a positive activator that is essential for transcription of the *otc* genes. The positive activator would be titrated out by the introduction of the cognate regulatory sequence on high-copy-number vectors. To date, however, there is no genetic evidence to support this proposal.

This chapter describes the isolation and characterisation of a promoter region which directs transcription of a putative *otcC-otcZ-otrA* polycistronic message that extends "leftwards" (as the cluster is drawn) to the terminator at the 3'-end of *otrA* during oxytetracycline production.

A promoter can be defined as the region of DNA involved directly in the binding of RNA polymerase to initiate transcription. Although RNA polymerase interacts with DNA from approximately 50bp upstream (-50) to about 20bp downstream (+20) of the transcription start point, sequence-specific contacts are not found ordinarily downstream of the transcription start site (Simpson, 1982). In general, the promoters recognised by different forms of RNA polymerase holoenzyme ( $E\sigma$ ) have characteristic consensus sequences (Grossman and Losick, 1986). Although the conserved sequences for each promoter class differ from each other, two hexameric regions of nucleotide conservation, which precede the transcription start site, are found typically in each promoter class. The conserved hexameric regions of the consensus sequence for the major class of eubacterial promoters, which are recognised by  $E\sigma^{70}$  in *E. coli* and  $E\sigma^{43}$  in *B. subtilis*, are centred approximately 35bp and 10bp preceding the transcription start site and

are separated from each other by 16-18bp (Hawley and McClure, 1983; Moran *et al.*, 1982). These hexamers are referred to as the "-35" (TTGACA) and "-10" (TATAAT) regions respectively. Mutational analysis and chemical-protection experiments have established that promoter recognition is largely (although not exclusive) determined by base-specific contacts within these two conserved hexameric regions of the promoter sequence (Simpson *et al.*, 1982; Grossman and Losick, 1986).

A consensus sequence has been derived for ten promoters from *Streptomyces*, which resemble the major class of eubacterial promoters (Hopwood *et al.*, 1986). In the *Streptomyces* consensus, the sequence of the -35 region and the spacing of the conserved hexameric regions are identical to the consensus for the major class of eubacterial promoters. However, the -10 region of the *Streptomyces* consensus (TAGGAT) differs slightly from that of *E. coli* and *B. subtilis*. This disparity could be a consequence of one or more of the following.

(i) In *E. coli* and *B. subtilis* a single abundant  $\sigma$  species directs transcription of the general "housekeeping" enzymes. In contrast, *S. coelicolor* contains a family of four distinct, but closely related,  $\sigma$  polypeptides that have high amino acid sequence identity to the major  $\sigma$  species found in *E. coli* ( $\sigma^{70}$ ) and *B. subtilis* ( $\sigma^{43}$ ) in the regions which are thought to make specific contacts with promoters (Takahashi *et al.*, 1988; Buttner *et al.*, 1990). It has been proposed that these  $\sigma$  factors recognise different classes of promoters with slight differences in the sequence or spacing of the -35 and -10 recognition sites (Westpheling and Brawner, 1989). It is possible, therefore, that the *Streptomyces* consensus (Hopwood *et al.*, 1986) was derived from four classes of promoters. Only one of these classes may be recognised by a  $\sigma$  factor that is the equivalent of the major sigma factor found in *E. coli* or *B. subtilis*.

(ii) The actual consensus sequence of promoters which direct the transcription of general "housekeeping" genes in *Streptomyces* may be slightly different due to the significant diversion in the G+C composition of *Streptomyces* compared with *E. coli* or *B. subtilis*.

(iii) The sample of promoters used to construct the *Streptomyces* consensus may not have been representative.



### 5.1.1 Overall Strategy

The results presented in Chapters 3 and 4 suggested that *otrA*, *otcZ* and possibly *otcC* were transcribed as part of a polycistronic message during antibiotic production. Transcriptional and functional analysis of the gene clusters for actinorhodin (Malpartida and Hopwood, 1986), methylenomycin (Chater and Bruton, 1985) and tylosin (Hershberger, 1989) has indicated that many of the genes within these respective cluster are organised in large transcription units extending up to 9.5kb (for further discussion, see Section 1.5). The *otcC*, *otcZ* and *otrA* genes are contained within a relatively short region of DNA of approximately 5.5kb. As the transcription and genetic organisation of the *otc* cluster upstream of *otcC* has not been studied intensively, it was considered possible that other genes could be co-transcribed with *otcC*, *otcZ* and *otrA* during antibiotic production. It was calculated, however, that there was a reasonable probability that the promoter(s), which directed transcription of *otrA*, *otcZ* and possibly *otcC* would be located within the 5.5 kb region that extends from the *Pst*II<sub>8</sub> site, which is internal to *otcC*, to the *Pst*II<sub>18</sub> site that is located approximately 4.5 kb upstream of the 5'-end of *otcC*.

As transcripts had been shown to terminate at the 3' end of *otrA* (Doyle, 1987), perhaps a straightforward approach to delineate the location of the promoter(s) directing transcription of *otrA* during antibiotic production would have been to blot total cellular RNA to nitrocellulose (Alwine *et al.*, 1977; Fourny *et al.*, 1988) and hybridize with radio-labelled probe derived from *otrA*. However, Northern analysis has never been used successfully within the laboratory at Glasgow (Alves, 1989; Doyle, pers. comm.). This strategy was not used, therefore, to locate the 5'-end of the polycistronic message. The overall approach that was taken during this work, was to screen for promoter activity in a library of *Sau*3AI fragments from the 5.5kb *Pst*II<sub>8</sub>-18 region, which had been constructed in the promoter-probe vector pIJ2843 (Clayton and Bibb, 1990), and then to locate the promoters in the *otc* cluster by Southern analysis (see section 2.3.15).

The pIJ2843 vector system can be used to measure *in vivo* promoter activity in *Streptomyces* spp, including the *S. rimosus* production strains. Neither the pIJ486/7 nor the pARC1 systems, which were described previously, can be used in *S. rimosus* for the reasons discussed in Section 3.1.2. The pIJ2843 vector is based on the low-copy-number SCP2\* replicon and utilises the *xylE* gene from *Pseudomonas putida* as the reporter of transcription (Clayton and Bibb, 1990). The product of the *xylE* gene is a catechol dioxygenase, which converts the colourless substrate catechol to an intensely yellow hydromuconic semi-aldehyde (Zukowski *et al.*, 1983). Catechol dioxygenase activity can be measured using a simple colourimetric assay. Using a similar vector, the activity of *galp1-xylE* fusions in crude extracts of *S. coelicolor* A3(2) was shown to reflect faithfully the intracellular levels of RNA, as determined by quantitative dot blots (Ingram *et al.*, 1989). Several useful features have been incorporated into the pIJ2843 vector. A multiple cloning site (MCS) upstream of the promoterless *xylE* gene assists the insertion and excision of fragments of DNA to be tested. The incorporation of a terminator from the coliphage fd upstream of the MCS prevents transcripts from vector promoters extending to the reporter gene. Translational stop codons have been created between the MCS and the *xylE* reporter gene in all three forward frames to prevent translational fusions which could affect either the expression of the *xylE* gene or the stability and/or specific activity of the *xylE* gene product. The terminator to from the coli phage lambda has also been inserted downstream of the reporter gene to prevent transcriptional read-through from the "test" promoter affecting the expression of genes in vector sequences and thus, possibly vector stability.

## 5.2 Results and Discussion

### 5.2.1 Construction of a Promoter Library In pIJ2843

The pIJ2843 vector was digested with *Bam*HI and treated with CIAP to remove the 5'-phosphate from the cohesive ends generated by linearisation of the plasmid (Section 2.3.7). To verify the efficiency of dephosphorylation and to ensure that the vector had retained its ability to ligate with "insert" DNA, the following ligation reactions were set up;

(i) pIJ2843 (100ng) treated with *Bam*HI and CIAP plus the 221bp *Bcl*II-*Bam*HI<sub>6</sub> fragment (3.5ng) containing *otrAp1*, (ii) pIJ2843 (100ng) treated with *Bam*HI and CIAP and (iii) as (ii) but not incubated with T4 ligase. The molar ratio of insert to vector in the test ligation (pIJ2843 plus "insert" DNA) was approximately 2:1.

Samples ( $1/10^6$ ) of the above ligation mixtures were introduced to *S. lividans* TK24 by genetic transformation. Serial dilutions of the transformed protoplasts were then incubated on regeneration plates to determine the efficiencies of transformation (Table 5.1).

Table 5.1 Assay of the Efficiency of Phosphatase Treatment.

	Transformation efficiency ( <u>transformants.<math>\mu</math>g<sup>-1</sup> vector</u> )
Supercoiled pIJ2843	$2.52 \times 10^7$
Dephosphorylated vector	$1.83 \times 10^4$
Self-ligated dephosphorylated vector	$1.23 \times 10^4$
Dephosphorylated vector ligated to test "insert" (221bp <i>Bcl</i> II- <i>Bam</i> HI fragment)	$3.17 \times 10^5$

The low number of transformants recovered after self-ligation of the linearised vector, which had been treated with CIAP, suggested that the 5'-phosphate groups had been removed with reasonable efficiency. As the transformation efficiency of the dephosphorylated vector, which had been ligated with the test "insert", was at least an order of magnitude greater than the control ligations, the linearised vector appeared to be able to ligate with "insert" DNA. The similar transformation efficiencies of 5'-dephosphorylated vector, which had been incubated with and without ligase, indicated that the background transformants were due primarily to incomplete digestion of the vector with *Bam*HI (only 0.5ng of undigested plasmid would have been required to produce  $10^4$  transformants). Comparison of the transformation efficiencies suggested that greater than 95% of the transformants from the test ligation should contain recombinant molecules.

Using sequence data for over 8kb of the oxytetracycline cluster, it was determined that *Sau3AI* cleaves streptomycete DNA on average every 160bp. As restriction sites for *Sau3AI* occur frequently, it was possible that these sites might occur in sequences required for transcription. To increase the probability of isolating promoters and any cognate regulatory sequences on the same fragment, the 5.5kb *PstI* fragment from pPFZ105 (a Pfizer construct) was only digested partially with *Sau3AI*.

Conditions were established for the partial cleavage of the 5.5kb *PstI* fragment with *Sau3AI* by adding a fixed amount of enzyme and monitoring the progress of the reaction. In a pilot experiment, 500ng of the 5.5kb *PstI* fragment was resuspended in 29 $\mu$ l of 1X *Sau3AI* restriction buffer (as supplied by BRL) and equilibrated to 37°C for 3-5min, before the addition of 0.5 units of *Sau3AI*. After incubation for 5, 15, 30, 45 and 60mins, 6 $\mu$ l aliquots were removed, added immediately to 4 $\mu$ l of 5X AGL buffer, incubated at 70°C for 15min to heat inactivate the restriction endonuclease and placed on ice until all the samples had been processed. The reaction products were then separated by electrophoresis on a 6% (w/v) polyacrylamide gel and compared with 100ng of the *PstI* fragment which had been digested completely with *Sau3AI*. Using this system, products that were digested completely (15-800bp) were resolved easily (results not shown).

Digestion of the 5.5kb *PstI* fragment was almost complete after incubation for 60min. After incubation for 45min, the fluorescence (as a result of ethidium bromide staining) of bands corresponding to products which were completely digested was approximately half that of the corresponding fragments in the sample which had been digested with an excess of *Sau3AI*. Although products which were completely digested could be seen in the sample digested for 30min, the majority of the fluorescence came from fragments greater than 900bp (result not shown). As the regulation of prokaryotic gene expression involves typically the binding of ancillary proteins to regulatory sites close to the promoter (usually less than 150bp from the transcription start site), it was anticipated that promoters and cognate regulatory sequences

could be isolated together on fragments smaller than 900bp. The 5.5kb *Pst*I fragment was eventually digested for 37min. All the products of complete digestion were represented and the majority of the digestion products were less than 900bp (result not shown).

For shotgun cloning, the recommended concentration of vector in the ligation reactions is 0.4–0.8 fmoles. $\mu$ l<sup>-1</sup> ("M13 Cloning/Dideoxy Sequencing" published by BRL). Correspondingly, the concentration of pIJ2843 used in the ligation reactions was 6ng. $\mu$ l<sup>-1</sup> (0.8 fmoles. $\mu$ l<sup>-1</sup>). Assuming that partial digestion produced fragments with an average length greater than 320bp, fewer than 16 *Sau*3AI fragments should have been produced, on average, from a single *Pst*I fragment. To suppress the formation of concatemers, which would subsequently hinder the positioning of promoters using Southern analysis, the molar ratio of insert-to-vector in the ligations was <1:1. It was calculated that 3ng. $\mu$ l<sup>-1</sup> of the products from the partial digestion of the 5.5kb *Pst*I fragment would corresponded to a concentration of no greater than 0.8 fmoles. $\mu$ l<sup>-1</sup>. The ligation reaction was set up, therefore, using 120ng of pIJ2843 (treated with *Bam*HI and CIAP) and 3ng of *Sau*3AI fragments. As a control, 120ng of dephosphorylated pIJ2843 was ligated to itself.

Aliquots (1/10th) of the ligation reactions were introduced to *S. lividans* TK54 protoplasts, which could be transformed by supercoiled pIJ2843 with an efficiency of  $3.97 \times 10^7$  transformants. $\mu$ g<sup>-1</sup> of pIJ2843. From the test ligation (pIJ2843 and the 221bp *Bcl*II-*Bam*HI fragment), it was predicted that 12ng of dephosphorylated vector ligated to the partially digested *Sau*3AI fragments would produce approximately  $4 \times 10^3$  transformants. Aliquots (10X) containing 1/120 dilutions of the transformed protoplasts were incubated on regeneration plates. As expected, approximately 50 transformants grew on each plate ( $5 \times 10^5$  transformants. $\mu$ g<sup>-1</sup> DNA). The transformation efficiency for self-ligated dephosphorylated vector was  $1.56 \times 10^4$ . $\mu$ g<sup>-1</sup>. Consideration of these transformation efficiencies suggested that greater than 95% of the transformants would contain recombinant vector molecules.

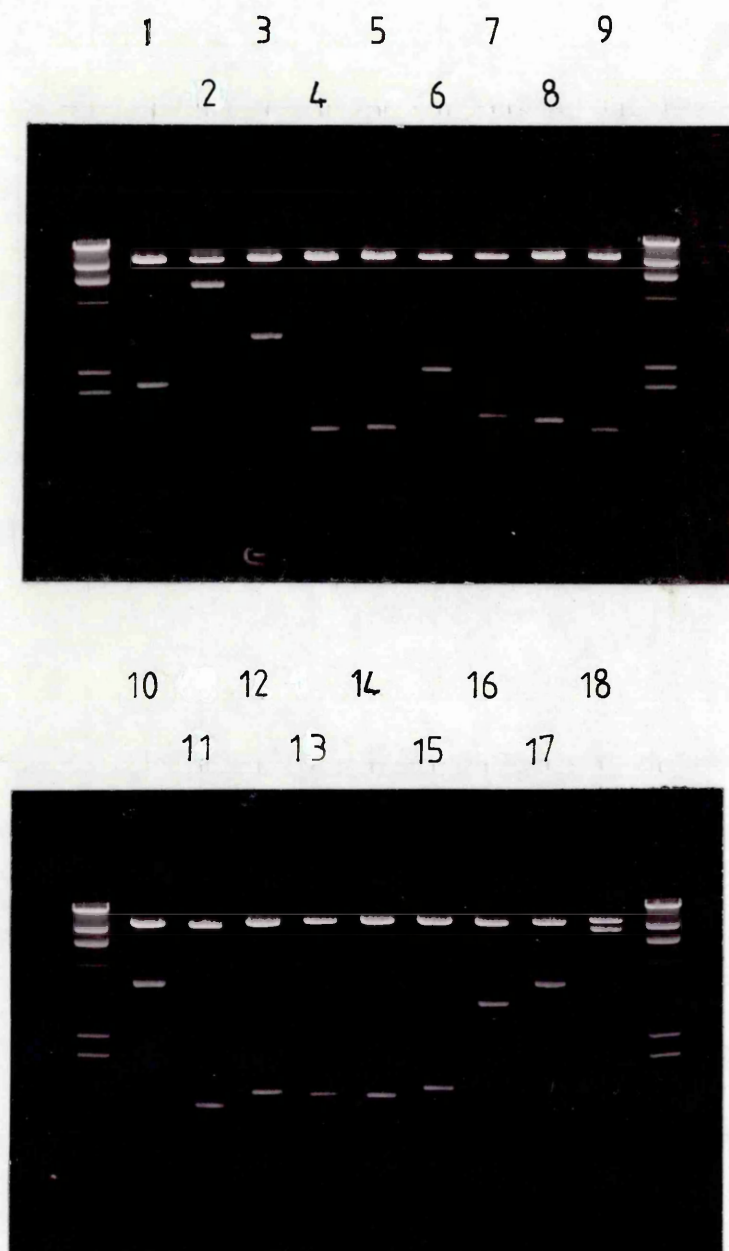
### 5.2.2 Assay For *in Vivo* Promoter Activity

The promoter responsible for the transcription of *otrA*, *otcZ* and possibly *otcC* has a distinct temporal activity associated with the slowing of growth rate and the onset of oxytetracycline (OTC) production (see Section 3.2.4). On solid media (Regeneration agar and Soya Mannitol agar), the production of OTC by *S. rimosus* is coincident with morphological differentiation. The demonstration that OTC can be produced by *S. lividans* when the entire cluster is cloned on a low-copy-number plasmid suggests that this heterologous host recognises the promoters responsible for the transcription of the OTC cluster (Binnie *et al.*, 1989). It is not known, however, if the production of OTC by *S. lividans* follows the same temporal pattern established for *S. rimosus*. Assays for *in vivo* promoter activity were carried out, therefore, after incubation of the transformants on regeneration plates (R2 medium) for 3, 4 and 5 days at 30°C. The regeneration plates were sprayed with an aerosol of 0.5M pyrocatechol (supplied in crystalline form by Sigma) and incubated at room temperature for 30min. For each time point, six well-isolated *xylE*<sup>+</sup> transformants with a range of pigmentation were selected for further analysis. Approximately 10% of the transformants were *xylE*<sup>+</sup>.

Secondary screening of the 18 isolates for catechol dioxygenase activity after growth on Regeneration agar and Hopwood's Minimum agar (both containing thiostrepton, 25µg.ml<sup>-1</sup>) for 3, 4 and 5 days failed to detect any catechol dioxygenase activity in isolate 5. All the other isolates were highly pigmented at each of the time points. *S. lividans* TK54 transformed with pIJ2843 did not exhibit any catechol dioxygenase activity.

### 5.2.3 Characterisation of the *xylE*<sup>+</sup> Transformants

Digestion of plasmid DNA from each of the 18 *S. lividans* transformants with *Hind*III and *Eco*RI released the "insert" fragment along with the *xylE* reporter gene from the rest of the vector. The fragments were resolved on a 0.8% (w/v) agarose gel (Figure 5.1) and then transferred to a Amersham Hybond-N<sup>TM</sup> membrane by capillary blotting (Southern, 1975) under alkaline conditions (see Section 2.3.15).



**Figure 5.1.** *Hind*III/*Eco*RI digests of independent isolates of the pIJ2843 constructs. The number above the lanes designates the isolate which was the source of the plasmid. Lambda *Hind*III markers were used to size the *Sau*3AI inserts.

Cleavage of the recombinant vector molecules with *Sph*I and *Pst*I indicated that gross rearrangements of the vector had not occurred (result not shown). The recombinant plasmids contained inserts ranging in size from less than 300bp to approximately 9.0 kb in isolate 18. The most likely explanation for the large insert present in this isolate is that a concatemer of large *Sau*3AI fragments has been cloned in this recombinant.

To localise the inserts containing promoter activity, digests of the 5.5kb *Pst*I fragment (*Sph*I, *Sst*I, *Sph*I/*Kpn*I, *Kpn*I and *Sau*3AI) were separated on a 0.8% (w/v) agarose gel and transferred to nylon membrane as described above. *Hind*III/*Eco*RI fragments from recombinant vectors with small *Sau*3AI inserts were selected preferentially for Southern analysis. The *Hind*III/*Eco*RI fragment from isolate 12, which contained the reporter gene and a *Sau*3AI insert of less than 300bp, was used to synthesize a radio-labelled probe by the "Random Primer" method (see Section 2.3.9.1). The probe hybridized to the 920bp *Sph*I<sub>12</sub>-*Sst*I<sub>13</sub> region (Figure 5.2, panel B). The *Sph*I-*Sst*I fragment was then isolated and used to prepare a radio-labelled probe. This probe hybridized with the plasmid DNA of only eight isolates (Table 5.2), indicating that more than one promoter element must exist in the 5.5kb region. To determine the minimum number of DNA elements within this region, which could direct the transcription of *xylE*, further rounds of hybridization were undertaken. Using the approach described above, the *Hind*III-*Eco*RI fragment from isolate 14, which did not hybridize with the *Sph*I-*Sst*I probe and contained a small *Sau*3AI insert (<300bp), was used to probe digests of the *Pst*I region. This probe hybridized to DNA from the 200bp *Kpn*I<sub>16</sub>-*Sst*I<sub>17</sub> region (Figure 5.2, panel C). In turn, the *Kpn*I-*Sst*I fragment was used to probe plasmid DNA from the 18 *xylE*<sup>+</sup> isolates. This probe hybridized to the plasmid DNA from seven *xylE*<sup>+</sup> transformants, including two which had been identified using the *Sph*I-*Sst*I probe (Table 5.2). As several of the transformants contained plasmids with large *Sau*3AI inserts (>1kb), it was not unexpected that certain plasmids would hybridize with more than one probe from the 5.5kb *Pst*I region. After this round of analysis, plasmid DNA from only five *xylE*<sup>+</sup> transformants had not hybridized with



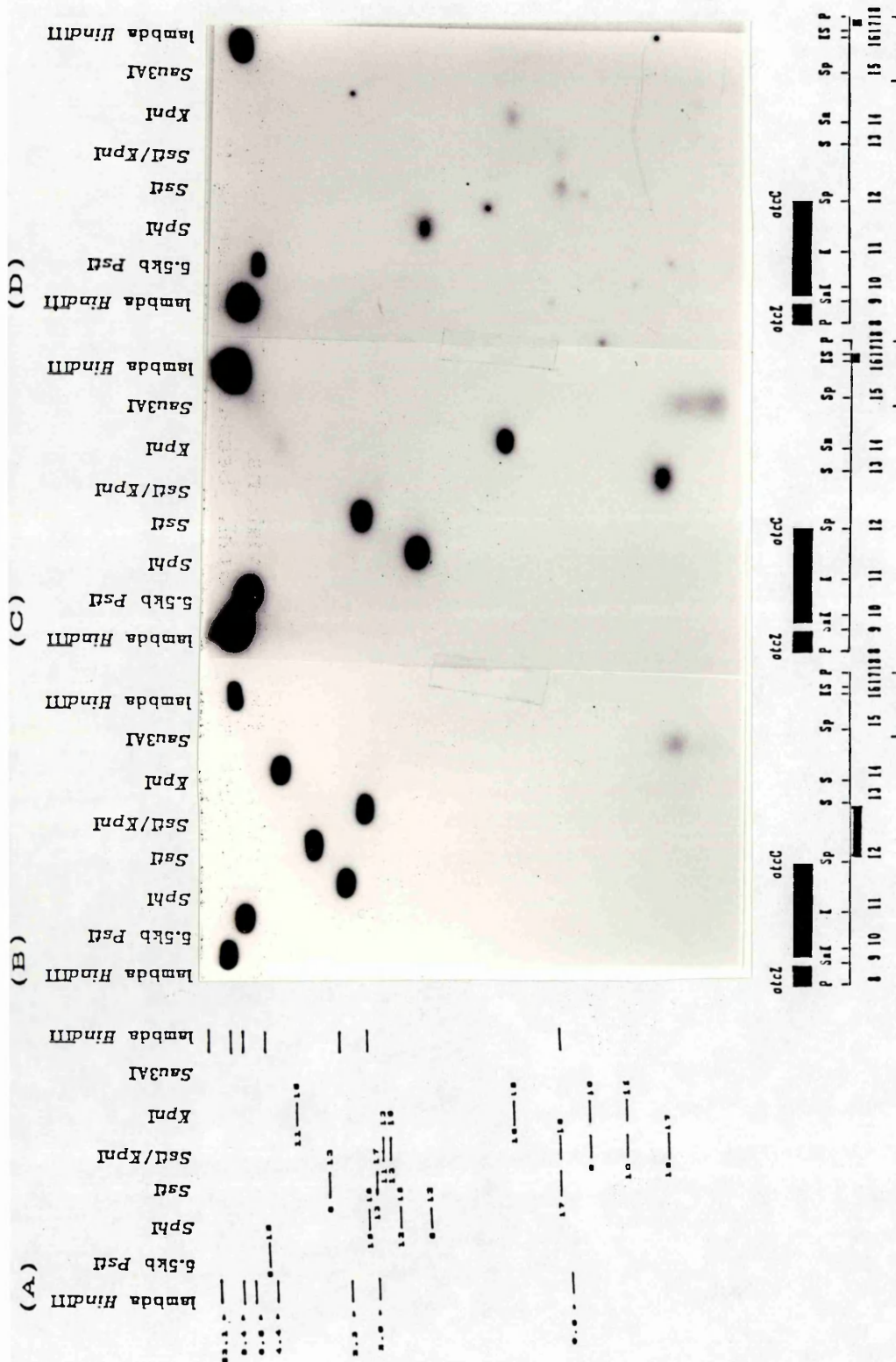
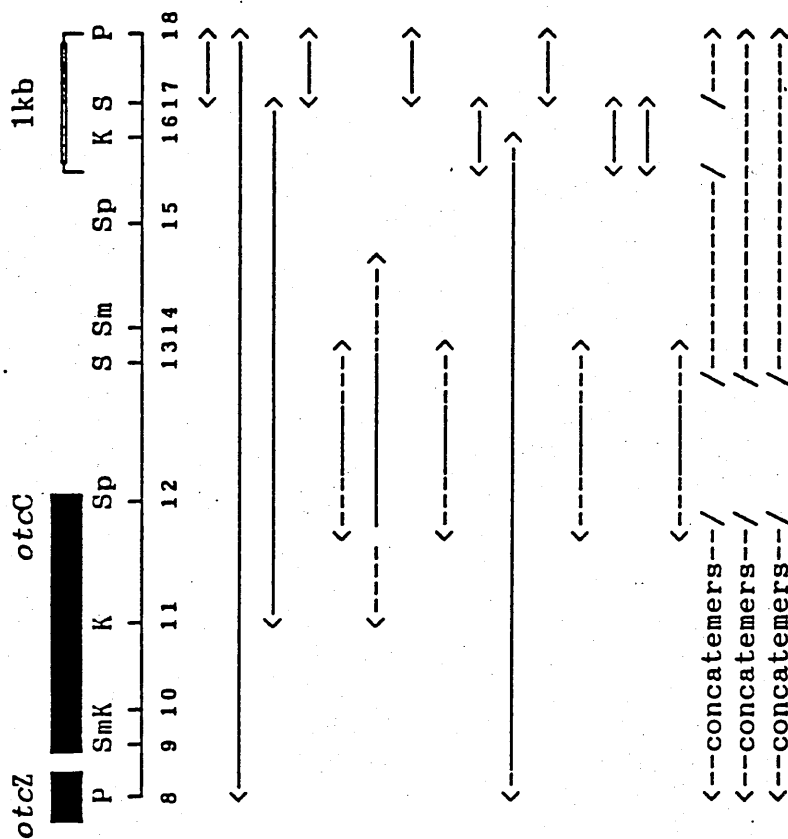


Figure 5.2. Location of the *Sau3AI* fragments, which were cloned in the xyle-positive transformants within the 5.5 kb *PstI* region. (A) Idealized representation of the fragments that were resolved by agarose gel electrophoresis and transferred to a Hybond- $N^{+}$  membrane. The numbers indicate the restriction sites used to derive the fragments. (B) Autoradiograph of the Hybond- $N^{+}$  membrane after hybridization with

radio-labelled probe derived from the insert in the pIJ2843-based construct of isolate 12. (C) and (D) as (B) except hybridized with probe derived from isolates 14 and 18 respectively. The location of the *Sau3AI* insert within the 5.5 kb *PstI* region is indicated below each autoradiograph. Abbreviations: K, *KpnI*; P, *PstI*; S, *SacI*; Sm, *SmaI*; Sp, *SphI*.



Isolate	Approx. Insert Size (kb)	Probe Hybridizations			
		<i>SphI</i> / <i>SstI</i>	<i>KpnI</i> / <i>SstI</i>	<i>SstI</i> / <i>PstI</i>	
1	0.4	-	-	-	+
2	5.0	+	+	+	+
3	3.5	+	+	+	-
4	<0.3	-	-	-	+
5*	<0.3	+	-	-	-
6	1.0	+	-	-	-
7	<0.3	-	-	-	+
8	<0.3	+	-	-	-
9	<0.3	-	+	-	-
10	4.0	+	-	-	-
11	<0.3	-	-	+	+
12	<0.3	+	-	-	-
13	<0.3	-	+	-	-
14	<0.3	-	+	+	-
15	<0.3	+	-	-	-
16	3.5	-	-	-	+
17	4.0	-	+	+	+
18	>9.0	-	+	+	+

Table 5.1. Summary of the hybridizations between the plasmid DNA in the xyle-positive transformants and probes derived from fragments of the 5.5kb *PstI* region. The line (<---->) indicates the probable location of the *Sau3AI* fragments, which were cloned in the transformants, within the *PstI* region.

Abbreviations: K, *KpnI*; P, *PstI*; S, *SacI*; Sm, *SmaI*; Sp, *SphI*. \*Secondary screening indicated that isolate 5 was xyle-negative. *cHindIII/EcoRI* fragments, which contained the *Sau3AI* insert, from the plasmid DNA of these isolates were used originally to probe the 5.5kb *PstI* region.

either of the probes. A radio-labelled probe was synthesized, therefore, from isolate 11, one of the remaining isolates that contained a small *Sau3AI* insert. This probe hybridized to DNA from the 400bp *Sst*I<sub>17</sub>-*Pst*I<sub>18</sub> region (Figure 5.2, panel D). In turn, the *Sst*I-*Pst*I fragment was used to probe plasmid DNA from the 18 *xylE*<sup>+</sup> isolates. The *Sst*I-*Pst*I probe hybridized to all of the remaining *xylE*<sup>+</sup> isolates and several which had been identified previously (Table 5.2).

From the hybridization data, it was possible to deduce the most probable locations of the *Sau3AI* fragments, which were cloned in the *xylE*<sup>+</sup> transformants (including isolate 5) within the 5.5kb *Pst*I region (Table 5.2). It was necessary, however, to postulate that several of the recombinant vectors contained large concatemers of *Sau3AI* fragments in order to reconcile the hybridization data with the size estimates of the *Sau3AI* inserts.

The hybridization of three *Sau3AI* inserts from different transformants to three discrete areas within the 5.5kb *Pst*I region indicated that a minimum of three DNA elements within the *Pst*I region could promote expression of *xylE* in the pIJ2843 constructs. As different promoters could be directing transcription in recombinant plasmids that hybridize with the same probe, more than three DNA elements from the *Pst*I<sub>8-17</sub> region could be directing the transcription of *xylE* in the pIJ2843 constructs.

#### 5.2.4 Location and Characterisation of the Promoter Region Responsible for the Transcription of *otrA* During Antibiotic Production

The main objective of shotgun cloning *Sau3AI* fragments into pIJ2843 was to delineate the location of the promoter(s) responsible for the transcription of *otcC*, *otcZ* and *otrA* during antibiotic production. The location of promoter activity to the *Sph*I<sub>12</sub>-*Sst*I<sub>13</sub> fragment suggested that such a promoter(s) would be located between the 5'-end of *otcC* and the *Sst*I<sub>13</sub> site. Therefore, further analysis of the *xylE*<sup>+</sup> transformants was not undertaken at this stage.

The following is the rationale for concluding that the *SphI*-*SstI* fragment contained a promoter capable of directing transcription of *otcC*, *otcZ* and *otrA*. It was considered improbable that the *SphI*-*SstI* region could contain only promoter activity capable of directing transcription in the opposite direction to *otcC*. The mRNAs of divergent non-overlapping genes within other antibiotic clusters do not generally overlap to any great extent, for example; the erythromycin resistance gene (*ermE*) and an unidentified ORF of *Saccharopolyspora erythraea* (Bibb *et al.*, 1986; Bibb and Janssen, 1987) and the *tylH,D* region and *tylF* region of *Streptomyces fradiae* (Hershberger, 1989). The organisation of transcription units within antibiotic clusters is discussed in detail within Section 1.5. Should the *SphI*-*SstI* region contain bidirectional non-overlapping open reading frames, it would be highly likely that the promoters responsible for their transcription would be located between the 5'-end of *otcC* and the *SstI* site.

The strong codon bias of *Streptomyces* genes (Bibb *et al.*, 1984) would appear to place severe restrictions on the coding capacity of the five reading frames which overlap protein coding regions. Only a couple of publications have presented tentative evidence for the presence of bidirectional, overlapping PPCRs in *Streptomyces*; ORF85 and ORF79 in the plasmid pIJ101 (Kendall and Cohen, 1988) and ORFs 1215 and 330 in the Insertion Element IS110 (Bruton and Chater, 1987). Interestingly, both of these examples are derived from extrachromosomal elements, which might be the subject of strong selective pressure to minimise the amount of DNA required to encode essential functions. To date, bidirectional PPCRs which overlap extensively have not been found within the gene clusters for antibiotics. It was considered unlikely, therefore, that a divergent mRNA originating within the *SphI*-*SstI* region would overlap with a large polycistronic message that originated upstream of the *SstI*<sub>13</sub> site and extended to *otrA*. Following this line of argument, it was concluded that a promoter(s) located between the 5'-end of *otcC* and the *SstI*<sub>13</sub> site should direct transcription towards *otcC*. It was also considered possible that another convergent gene(s) within the *SphI*-*SstI* region could precede *otcC* in the putative transcription unit that extends to the 3'-end of *otrA*.

To corroborate, the above deductions and to facilitate the orientation and location of the *Sau3AI* inserts in the pIJ2843 constructs, the *SphI*<sub>12</sub>-*SstI*<sub>13</sub> region was sequenced. Detailed analysis of the sequence is described in the following chapter. However, a schematic representation of the genetic organisation in this region is shown in Figure 5.3. The *SphI*-*SstI* region contains two potential protein coding regions (PPCRs) which are encoded on the opposite strand to *otcC*, *otcZ* and *otrA*. The predicted AUG start codon for *otcX* PPCR1 is 120bp to the right of the *SphI* site. As the translation initiation codon for *otcC* is 35bp to the left of the *SphI*<sub>12</sub> site, the non-coding region between *otcC* and *otcX* PPCR1 is only 155bp. The recognition sequences for promoters are not found generally within protein coding regions. Therefore, it was considered probable that the promoter for the putative *otcC-otcZ-otrA* polycistronic message would be contained within the intergenic region between *otcC* and *otcX* PPCR1.

#### 5.2.4.1 Location of the 5' Termini of the Divergent Transcripts

Primer extensions were used to map and quantitate the 5'-termini of the putative *otcC-otcZ-otrA* polycistronic message and the divergent transcript from the opposite strand. The technique involves hybridizing the RNA of interest to an excess of a synthetic oligonucleotide, which is radio-labelled at the 5'-end. Reverse transcriptase is then used to extend the primer to produce DNA complementary to the RNA (cDNA). The same oligonucleotide can also be used to generate dideoxy-mediated sequencing ladders (Sanger *et al.*, 1977), which can be used to size the cDNA. Using this approach, it is essential that the oligonucleotide primers bind to a site within 50-150 nucleotides of the 5'-terminus of the mRNA. Primers that hybridize to more-distal sites generate frequently heterologous extension products due to the tendency of reverse transcriptase to pause in regions of high secondary structure within the RNA. This is a particularly important consideration when analysing streptomycete RNA, which contains a preponderance of inverted repeats due to the high G+C content (Enquist and Bradley, 1971).

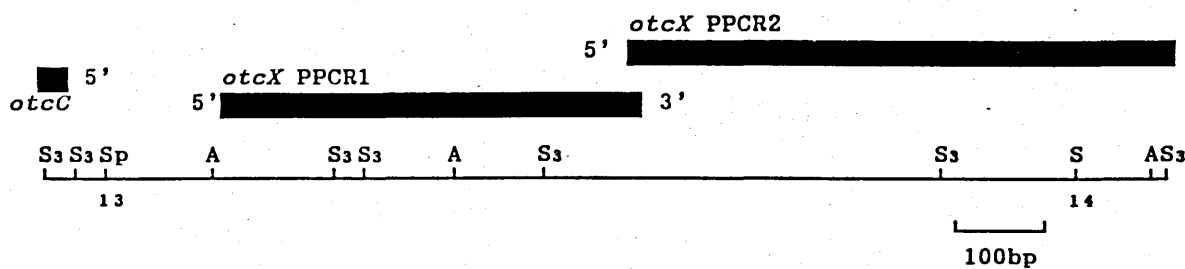


Figure 5.3. Schematic representation of the genetic organisation within the *Sph*I<sub>12</sub>-*Sst*I<sub>13</sub> region. The closed boxes (■) indicate the positions of the PCRs. Abbreviations: A, *Ava*I; S, *Sac*I; Sp, *Sph*I; S3, *Sau*3A1. Modified from Figure 6.3.

Reverse transcriptase is available from two sources, Molony Murine Leukemia Virus (M-MuLV) and Avian Myeloblastosis Virus (AMV), which have optimum temperatures of 37°C and 42°C respectively. Initial attempts at primer extension using M-MuLV reverse transcriptase were unsuccessful. Although a major extension product could be identified, a heterogeneous population of extension products was generated (result not shown). Under the conditions recommended by Maniatis *et al.* (1989), MuLV reverse transcriptase appeared to have considerable problems extending through regions of streptomycete RNA.

Homogeneous extension products were obtained, however, using a protocol communicated by L.V. Wray, Jr (Boston, University), which used AMV reverse transcriptase to extend the primer. The decreased stability of secondary structure in RNA at the higher incubation temperature used with AMV reverse transcriptase may explain the observed decrease in heterogeneous extension products.

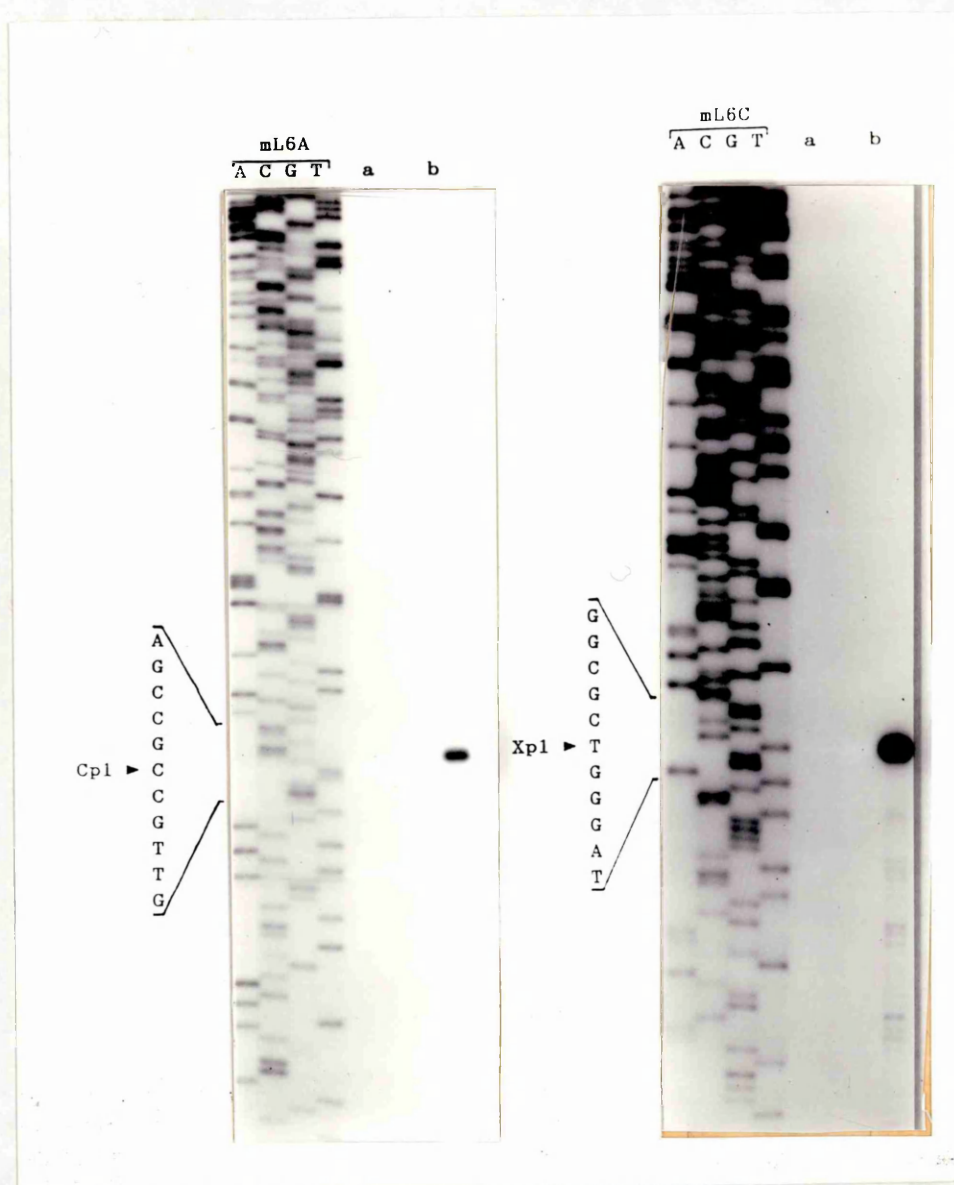
The 5'-termini of transcripts for several genes associated with secondary metabolism are also the first nucleotide of the proposed translation initiation codons, for example, the 23S rRNA methylase gene from *S. erythraeus* (Bibb *et al.*, 1985), the streptothricin acetyl transferase (*sta*) gene from *S. lavendulae* (Horinouchi *et al.*, 1987), the aminoglycoside phosphotransferase (*aph*) gene from *S. fradiae* (Janssen *et al.*, 1989) and the gene (*afsA*) responsible for A-factor biosynthesis in *S. griseus* (Horinouchi *et al.*, 1989). Oligonucleotide primers, C68 and X84, were synthesized, therefore, that hybridized to sites in the RNA 68 and 84 nucleotides downstream from the translation initiation codons of *otcC* and *otcX* PCR1 respectively (Figure 5.6). Two M13mp18-based templates mL6A and mL6C (Butler, unpublished), which differed only in the orientation of their inserts, were used to generate sequencing ladders for sizing the cDNA which were synthesized from the divergent transcripts. Both templates contained the 2.05 kb *Sma*I<sub>9-14</sub> fragment; the *Sma*I<sub>14</sub> site being nearest the binding site for the universal primer (UP) in mL6A (for restriction sites, see Figure 5.10). The oligonucleotides, C64 and X84, were also used to generate sequencing ladders from mL6A and mL6C respectively.

Primer extension analysis identified single 5'-termini for both the putative *otcC-otcZ-otrA* message and the divergent transcript in total RNA isolated from *S. rimosus* M15883 grown on TS1 for 46hrs (Figure 5.4, Panels A and B). The leader sequences of the divergent mRNA for *otcC* and *otcX* were 30 and 25 nucleotides respectively. The transcripts were separated by 100bp: they were non-overlapping (Figure 5.6).

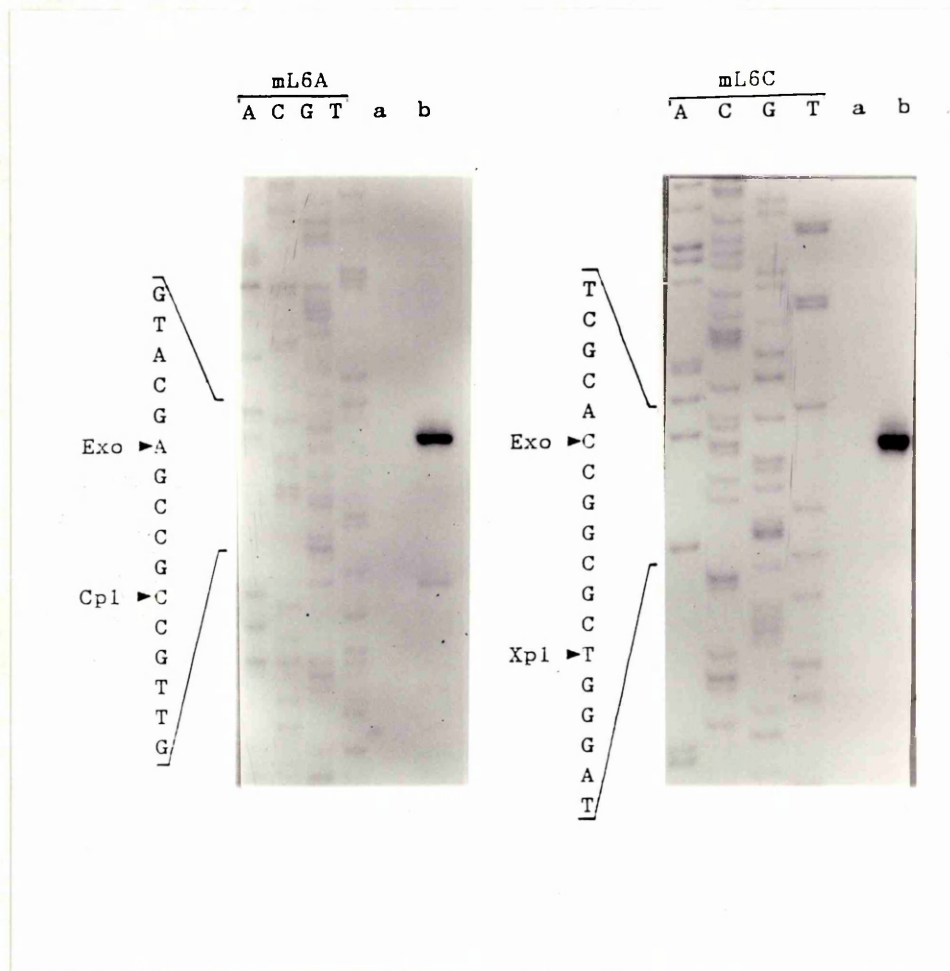
Analysis of hypothetical transcripts in the intergenic region between *otcC* and *otcX* PPCR1 using the FOLD program (Devereux *et al.*, 1984; see Section 3.2.4) did not identify any significant secondary structure which could possibly block the extension of the complementary DNA on hypothetical transcripts originating outwith this region. Exonuclease VII protection analysis of the *otcC/otcX* promoter region was carried out, however, to corroborate the primer extension data.

A single-stranded probe (220nt), complementary to the sense strand of *otcC*, was generated from mL6A by extension from oligonucleotide C64 (which was labelled at the 5' end) and digestion of the product with *Ava*I (see Section 2.3.18). Similarly, single-stranded probe (199nt), complementary to the sense strand of *otcX*, was produced by extension from 5'-labelled oligonucleotide X84, which was hybridized to mL6C, and digestion of the product with *Sph*I. As these particular probes do not contain sequences derived from coli phage M13mp18 at their 5'-termini, the position in the sequence ladder with which they co-migrate corresponds to the actual transcription start site. Exonuclease VII analysis (Figure 5.5) revealed the same transcription start points observed by primer extension, indicating that the extension were not interrupted by secondary structure in the mRNA. As Exonuclease VII cannot digest DNA to a flush end (Rose and Botstein, 1983), the major Exonuclease VII-resistant fragments migrated with an apparent size 5-7nt larger than the cDNA products of primer extension.





**Figure 5.4.** Primer extension analysis of the 5' ends of the putative *otcC-otcZ-otrA* message and the divergent transcript. (A) The extension product from oligonucleotide C68 hybridized with 10 $\mu$ g of total RNA isolated from *S. rimosus* M15883 grown on TS1 for 46hrs (see Figures 3.6 and 3.7). (B) as (A) except extensions were from oligo. X84. The arrows (>) indicate the transcription start points of *otcCp1* and *otcXp1* in the sequencing ladders derived from oligo. C68 hybridized to mL6A and oligo. X84 hybridized to mL6C respectively. No extension products were generated from total RNA isolated from *S. rimosus* M15883S.



**Figure 5.5.** Exonuclease VII analysis of the divergent transcripts. (A) Single-stranded probe ( $1.3 \times 10^7$  Cerenkov cpm.pmol $^{-1}$ ) complementary to the anti-sense strand of *otcC* was generated by extension from 5'-labelled oligo. C64, which was annealed to mL6A, using unlabelled nucleotides and digestion with *Ava*I (see Section 2.3.18). 0.3pmoles of probe was hybridized at 37°C with 10µg of total RNA isolated from *S. rimosus* M15883S (lane a) and *S. rimosus* M15883 grown on TS1 for 46hrs (lane b). (B) Single-stranded probe ( $0.9 \times 10^7$  Cerenkov cpm.pmol $^{-1}$ ) was synthesized from mL6C by extension from 5'-labelled oligo. X84 and digestion with *Sph*I. 0.2pmoles of probe was hybridized to the same selection of RNA as in (A).

#### 5.2.4.2 Quantification of the 5' Termini of the Divergent

##### Transcripts.

Primer extensions were used to quantitate the 5'-termini of the putative *otcC-otcZ-otrA* polycistronic message and the divergent transcript in RNA isolated from *S. rimosus* M15883 (see Figure 3.6) at various points during the fermentation of oxytetracycline. Apart from taking rigorous steps to ensure that the same amount of RNA (10µg) in each sample was hybridized to the oligonucleotide primer, the extension reactions were undertaken as described previously (Section 5.2.4.1).

The amount of cDNA synthesized from total RNA, which was isolated from cultures grown on TS1 for 46hrs, was approximately 4-4.5 fold greater than the amount produced from RNA isolated after 93hrs (Figure 5.7, Panel A). The 46hr sample was isolated early in the production phase when the concentration of OTC in the media was low, approximately 0.6mg.ml<sup>-1</sup>. The 93hr sample represented the mid-point in the fermentation when the concentration of OTC was 3.2mg.ml<sup>-1</sup>. The amount of cDNA synthesized from RNA isolated after 69hrs was slightly more than that produced from the 93hr sample. Previously, a single-stranded probe was used to measure quantitatively the amount of transcription between the 3' end of *otcZ* and the 5' end of *otrA* in the same RNA samples (Section 3.2.4). The relative amounts of single-stranded probe protected by transcripts within the different RNA samples (see Figure 3.7B) were similar to the amount of cDNA synthesized on the *otcC* transcript at the 5' end (Figure 5.7). These results were consistent with *otcC* being the first gene of a transcription unit that also includes *otcZ* and *otrA*.

Under similar fermentation conditions, the specific activity of anhydrotetracycline oxygenase, which is encoded by *otcC*, has been shown to rise rapidly during the transition from the period of intense growth to the stationary phase, reach a maximum level approximately 20 hours after OTC is first detected in the culture broth and then decline, within 20 hours, to a level seen at the onset of production (Butler, unpublished result). The changes in the specific activity of ATC oxygenase, which reflect approximately the changes in the abundance of

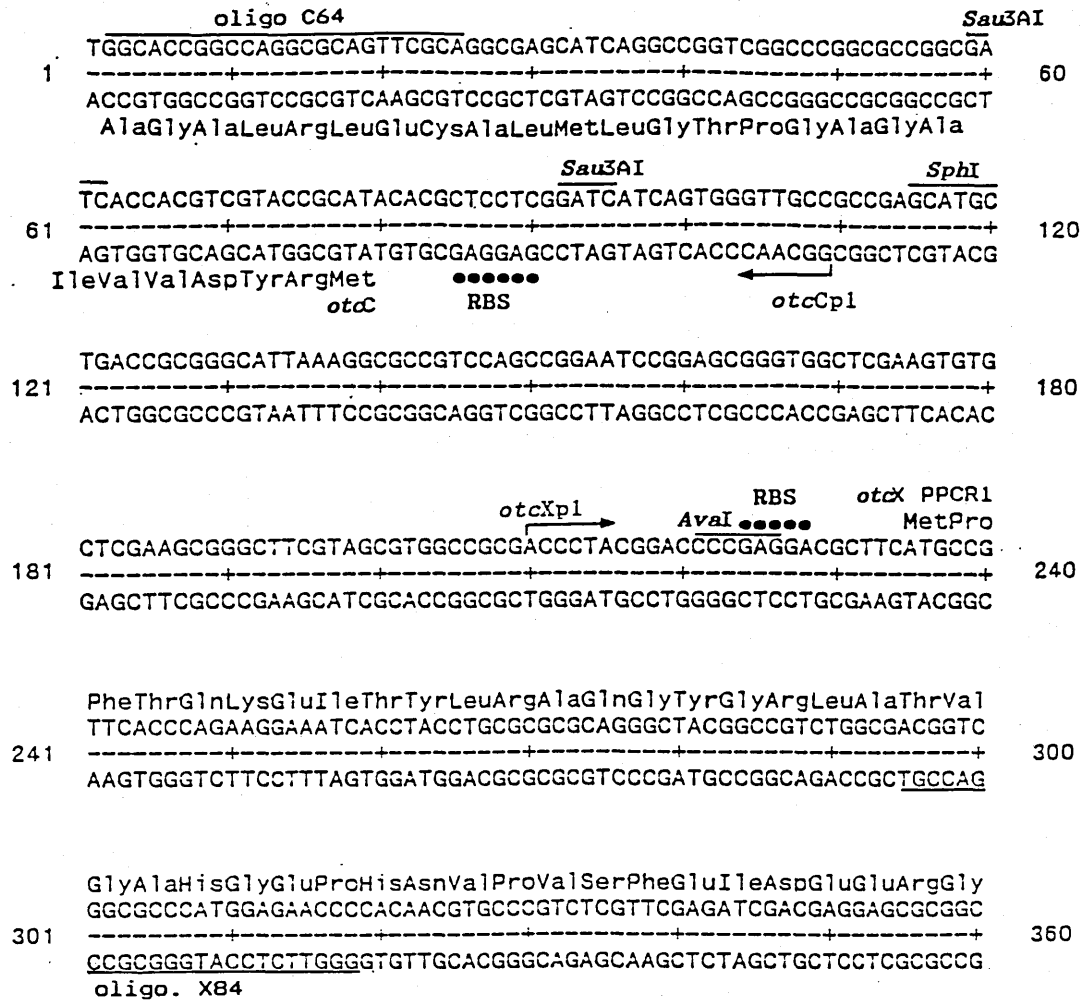
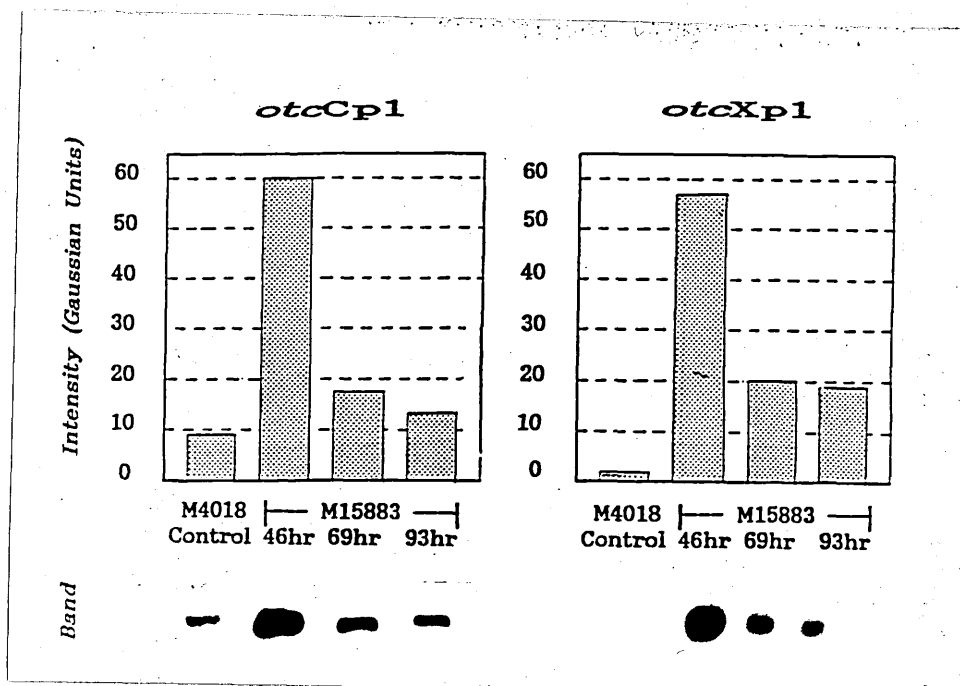


Figure 5.6. Annotated sequence of the *otcCp1/otcXp1* promoter region showing the position of the putative translation start codon, ribosome binding sites, transcription start sites and promoter sequences for *otcC* and *otcX*.



**Figure 5.7.** Quantitative measurement of cDNA synthesized from the divergent transcripts at the 5' end. (A) Oligonucleotide C64 (1ng), which was labelled at the 5' end, was annealed to 10µg of total RNA isolated from *S. rimosus* M4018 grown TSB, *S. rimosus* M15883 grown on TS1 for 46hrs, 69hrs and 93hrs (see Figure 3.6). (B) as (A) except oligo. X84 was annealed to the RNA samples. The primers were extended by AMV reverse transcriptase (see Section 2.3.19). The intensity of the images produced by autoradiography of the 6% (w/v) polyacrylamide gel used to resolve the cDNAs was determined by densitometry scanning (see Section 2.3.21).

the 5' end of the putative *otcC-otcZ-otrA* message, provided support for the proposal (Section 4.3) that the delay in the onset of oxytetracycline production can be attributed, at least in part, to the temporal regulation of transcription. It remains a formal possibility, however, that the catalytic activity of anhydrotetracycline oxygenase and possibly other enzymes involved in the production of OTC may also be regulated.

Changes in the abundance of the 5'-terminus of the divergent transcript followed a similar pattern to those determined for the putative *otcC-otcZ-otrA* polycistronic message (Figure 5.7, Panel B). Assuming that the primers, which were used to analyse the transcripts, were labelled with similar efficiency (no significant difference in the amount of radioactivity in the oligonucleotides was detected after removal of the unincorporated  $\gamma$ -<sup>32</sup>P ATP by column chromatography), the similarity in the intensity of the cDNA products produced by autoradiography suggests that in corresponding RNA samples the putative *otcC-otcZ-otrA* polycistronic message and divergent transcripts were present in similar amounts. The promoters for these transcripts would appear, therefore, to have similar temporal regulation.

Primer extension analysis using RNA isolated from cultures of *S. rimosus* M4018 grown on TSB, in which transcripts originating from *otrAp1* predominate (see Figure 3.7A) failed to detect any transcripts other than those originating from *otcCp1* and *otcXp1* (results not shown). It would appear, therefore, that single promoters direct transcription of the *otcC-otcZ-otrA* polycistronic message and the divergent transcript.

#### 5.2.5 Comparison of the *otcCp1/otcXp1* Promoter Sequences

Alignment of the DNA sequence from -50 to +1 of *otcCp1* and *otcXp1* revealed greater than 38% identity, without the need to insert gaps (Figure 5.8). Several other interesting features were also identified, which could be involved in the initiation and regulation of transcription.

```

      - - - > - - - >
GGATTCCGGCTGGACGGCGCCTTTAATGCCCGCGGTCAGCATGCTCGGCG  otcCp1
*      **** ** *      ** **      *** ** **
      - - - > - - - >
AGCGGGTGGCTCGAAGTGTGCTCGAAGCGGGCTTCGTAGCGTGGCCGCGA  otcXp1

      TTGaca                      tAGgaT      Streptomyces
                                      Consensus
      TTGACa                      TAtAaT      E. coli
                                      Consensus

```

Figure 5.8. Sequence comparisons of the *otcCp1* and *otcXp1* promoters. Transcription start sites and putative -10 and -35 regions are underlined. Asterisks (\*) indicate conserved nucleotides. Bold arrows (—>) indicate the location of the tandem repeats. The *E. coli* and *Streptomyces* consensus sequences shown are those of Hawley and McClure (1983) and Hopwood *et al.* (1986).

Prokaryote	Gene	Sequence	Homology
		5' - 3'	
<i>E. coli</i>	Consensus	———> ———> CTGTCATAAAACTGTCAC T      T  T      T	
<i>S. griseus</i>	<i>pabS</i>	CTGTCATGATACTGCGCG	66%
<i>S. lividans</i>	ORFX	CTTGACCTCAC.GTCACG	72%
<i>S. griseus</i>	<i>aphD</i>	CTGTGCGACATCTGGCAT	72%
		/C\	
<i>S. rimosus</i>	<i>otcC</i>	GGCTGGACGGCGCCTTTAAT	44%
<i>S. rimosus</i>	<i>otcX</i> PCR1	GGCTCGAAGTGTGCTCGAAG	33%

Figure 5.9. Alignment of possible regulatory sequences in phosphate-regulated promoters of *Streptomyces*. The consensus sequence for *pho* boxes in *E. coli* is that of Makino *et al.* (1990). The tandem repeat in the *E. coli pho* box is indicated by arrows (—>). The nucleotides in bold type are conserved with respect to the consensus *pho* box in *E. coli*. The alignment of the promoters for the *pabS*, ORFX and *aphD* genes was taken from Liras *et al.* (1990).

Centred around -10 there are hexamers in *otcCp1* (CAGCAT) and *otcXp1* (TAGCGT) which match reasonably well the -10 consensus sequence (TATAAT) for the major class of eubacterial promoters. The tentative -10 regions of these promoters show even higher similarity to the -10 region (TAGGAT) of the *Streptomyces* consensus (see Section 5.1) (Hopwood *et al.*, 1986).

An inspection of the promoter sequences 16-18bp upstream of the "-10" hexamers, revealed no similarity to the "-35" region of the consensus for the major class of eubacterial promoters (TTGACA). Centred around -36, however, hexamers in *otcCp1* (TGGACA) and *otcXp1* (TCGAAG), which are spaced 21bp upstream of the tentative -10 region, did show reasonable similarity to the consensus -35 region. Whilst, such similarities would be dismissed in *E. coli* due to the uncharacteristic distance between the putative recognition sequences (Mandecki and Reznikoff, 1982; Mulligan *et al.*, 1985; Stefano and Gralla, 1982), there is a precedent for this spacing in certain promoters from *B. subtilis*, which are regulated temporally. The *spoIIE* promoter (Guzman *et al.*, 1988) and *spoIIG* promoter (Kenney and Moran, 1987) both contain hexamers which show high similarity to the -35 and -10 regions of the consensus for the major class of eubacterial promoters. However, these regions have the atypical spacings of 21bp for the *spoIIE* promoter and 22bp for the *spoIIG* promoter respectively. Similarly, the spacing of the putative -35 and -10 regions of the *galp1* promoter from the *Streptomyces* galactose operon is unusually long (21 or 24bp) (Westpheling and Brawner, 1989).

For *galp1*, it was proposed that the major form of RNA polymerase could be made to recognise promoters with "-10" and "-35" regions that were separated by an atypical distance by regulatory proteins that interacted either with the promoter or with the RNA polymerase (Westpheling and Brawner, 1989). Another possibility is that one or more of the *Streptomyces*  $\sigma$ -like polypeptides encoded by the homologues of *rpoD* (the *hrd* genes- Takahashi *et al.*, 1988; Buttner *et al.*, 1990) could recognise promoters with sequences similar to the major eubacterial class, but separated by atypical distances (for discussion of



transcription initiation, see Section 1.9). It is possible, therefore, that the hexameric regions identified in *otcCp1* and *otcXp1* could be recognised by RNA polymerase. Transcription from *otcCp1* and *otcXp1* occurs predominantly during the early phase of antibiotic production. The unusual spacing of the putative -35 and -10 regions in these promoters may reflect this distinct temporal activity.

Between -41 and -23 of *otcXp1* a perfect tandem repeat of an octamer (GCTCGAA) was identified. The repeat sequences were separated by 3bp. Similarly, a hyphenated tandem repeat sequence was found in *otcCp1* (Figure 5.8). Direct repeats with a similar signature to those found in *otcCp1* and *otcXp1* are the recognition sites for transcription activators such as OmpR, ToxR and PhoB in *E. coli* (Mizuno and Mizushima, 1986; Miller *et al.*, 1987; Makino *et al.*, 1986).

To date, the best characterised of these regulatory proteins is PhoB, which positively regulates the phosphate regulon in *E. coli*, during phosphate limitation (see Section 1.7) (Makino *et al.*, 1986; Makino *et al.*, 1988). PhoB recognises and binds to a conserved regulatory sequence, designated the *pho* box, activating transcription (Makino *et al.*, 1988). The consensus sequence for the *pho* box consists of a tandem repeat of a heptamer (CTGTCAT) separated by 4bp. It has been suggested that each *pho* box may be bound by two PhoB subunits. Each PhoB subunit probably makes contact with the same side of the DNA molecule, as one turn of the helix of B-form DNA is 10.5bp long, similar to the distance between the centres of the repeated sequences (10bp). The *pho* boxes of *E. coli* promoters regulated by phosphate are located at the same position, 10bp upstream of the -10 regions. It has been proposed that PhoB protein bound to these *pho* boxes, substitutes for a typical -35 region in the phosphate-controlled promoters of *E. coli* (Makino *et al.*, 1986, 1988).

Similar to the *pho* boxes, the tandem repeats in the promoter sequences of *otcCp1* and *otcXp1* were found 10bp upstream of the tentative "-10" regions. The centres of these repeats were also separated by 10bp. It is conceivable, therefore, that the tandem repeats

between -41 and -23 of *otcCp1* and *otcXp1* may be recognition sites for a common regulatory protein that binds at the onset of oxytetracycline production activating transcription.

As the *otcCp1* promoter directs transcription of the ATC oxygenase, which has been shown to be regulated by phosphate (Behal 1979; Butler pers. comm.), the tandem repeats of *otcCp1* and *otcXp1* were aligned with the consensus for the *E. coli pho* boxes (Makino *et al.*, 1990) and *pho*-like sequences from the regulatory regions of *Streptomyces* promoters (Figure 5.9) (Liras *et al.*, 1990), which are also subject to phosphate control (see Section 1.7). The tandem repeats of *otcCp1* and *otcXp1* have 44% and 33% sequence identity to the *E. coli* consensus sequence respectively. Although the *pho*-like sequences of the *pabS*, *sphD* and ORFX promoters from *Streptomyces* have higher similarity to the *E. coli* consensus (>65%), the *otcCp1* and *otcXp1* matches are considered to be significant. The tandem repeats of both the *otc* promoters share the most highly conserved nucleotides of the *E. coli* consensus (CT---A-). It is proposed that the tandem repeats of *otcCp1* and *otcXp1* be referred to as *opl* (oxytetracycline *pho*-like) boxes.

The *S. rimosus* production strain M15883 was derived from a soil isolate, *S. rimosus* G7, through an extensive strain improvement programme, involving many rounds of mutagenesis and selection for survivors that produced higher titres of OTC. Transcription of the *otc* cluster may have been a target of strain improvement. It is possible, therefore, that any *pho* boxes in the oxytetracycline cluster may have diverged significantly from that of the original *S. rimosus* strain G7. This may explain why the tandem repeats of *otcCp1* and *otcXp1* show less homology to the *E. coli* consensus than the other *pho*-like sequences so far identified in *Streptomyces*. This possibility is currently being investigated by sequencing the *otcCp1/otcXp1* promoter region from *S. rimosus* G7, which has been amplified using the polymerase-chain reaction (Hunter, unpublished work).

The similarity of the *opl* boxes, which overlap with the putative -35 regions of the *otcCp1* and *otcXp1*, to the *pho* boxes in *E. coli* and the suppression of oxytetracycline production by high levels of phosphate, suggests that the onset of antibiotic production in *S. rimosus* may be controlled by a protein which under phosphate limitation activates transcription by binding to promoters within the *otc* cluster. This proposal could provide an explanation for the reduction in oxytetracycline production, which is observed when certain regions of the *otc* cluster are introduced into the production strains on high-copy-number plasmids, but not on low-copy-number plasmids (Butler *et al.*, 1989). Analysis of the DNA segments at the *otrA* end of the oxytetracycline cluster, revealed that all of the fragments containing *otcCp1* and *otcXp1* caused "switch-off" when they were cloned on the high-copy-number plasmids (Figure 5.10). It is possible that plasmid-borne copies of the tandem repeats in these promoters sequestered the putative regulatory protein, which is required for the transcription of the *otc* cluster. The activator protein may be the *Streptomyces*-equivalent of the PhoB protein found in *E. coli*.

An overview of the Pho regulon in *E. coli* is presented in the General Introduction (Section 1.7) and therefore, will not be described at length in the following discussion. The phosphate regulon in *E. coli* was characterised originally by analysing the expression of the alkaline phosphatase gene, *phoA* (for a lucid review and references, see Stock *et al.*, 1989). The *phoA* gene product hydrolyses phosphoesters to Pi in the periplasm and can be assayed easily by a simple plate test (Torriani and Rothman, 1961). In wild-type cells, the expression of *phoA* is suppressed by high levels of phosphate. Many of the loci involved in the uptake of phosphate were identified by characterising mutants, which expressed constitutively alkaline phosphatase. It would be interesting, therefore, to study the expression of alkaline phosphatase in *Streptomyces*. Should the *otc* genes be part of a phosphate regulon in *Streptomyces*, it would be predicted that mutations which cause the constitutive synthesis of the alkaline phosphatase at a high-level activity may also result in the constitutive production of OTC. Such mutants would demonstrate convincingly the integration of phosphate assimilation and OTC production.

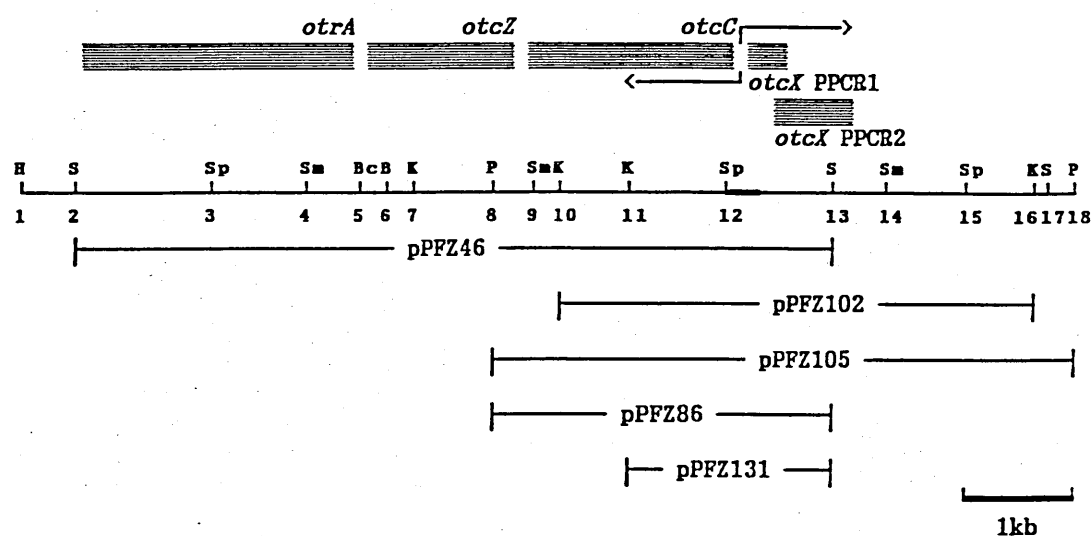


Figure 5.10. The "Switch-Off" phenomena at the "*otcA*" end of *otc* cluster. Lines (|—|) indicate the extent of the DNA segments which cause a significant reduction in the biosynthesis of oxytetracycline when introduced into the production strain, *S. rimosus*, on high-copy-number plasmids. A bold line (—) indicates the location of the *otcCp1/otcXp1* promoter region. Abbreviations: B, *Bam*HI; Bc, *Bcl*I; K, *Kpn*I; P, *Pst*I; S, *Sac*I; Sm, *Sma*I; Sp, *Sph*I.

The *opl* boxes may provide an alternative startpoint for investigating the mechanism(s) which controls the onset of oxytetracycline production in *S. rimosus*. By combination of sequencing and transcription analysis it should be possible to establish if *opl* boxes are a common feature of promoters in the oxytetracycline cluster. Should this be the case, then gel retardation assay using *opl* boxes as substrates may facilitate the purification of the putative response regulator OplB, that presumably activates transcription. The "reverse translation" of amino acid sequence from the amino-terminus of the purified protein, would allow an oligonucleotide to be generated, which should facilitate the isolation of the gene for putative regulator to be isolate from a DNA library of *S. rimosus* M15883.

Finally, it is interesting to speculate why the production of an antibiotic may be part of a phosphate regulon. On solid medium, the production of antibiotic often coincides with the switch to morphological development, when the aerial mycelium is envisaged to gain sustenance from the lysing substrate mycelium. It has been argued that the general role of antibiotic production is to prevent the invasion of the lysing substrate mycelium by microorganisms such as bacilli (Chater, 1984). The production of antibiotics does not always accompany the morphological development of *Streptomyces*, for example, *S. rimosus* produces OTC on L-agar plates, but does not sporulate, while sporulation but not antibiotic production occurs on plates containing tryptone soya agar (TSB, containing 2% {w/v} agar) (Hunter, pers. comm.). In soil, antibiotic production may not be necessarily a prelude to the morphological development of *Streptomyces*. When growth is limited by low levels of phosphate, the role of oxytetracycline may be to kill other microorganisms, resulting in the release of catabolites, including phosphate, which could be used to sustain the growth of the producing organism.

## CHAPTER 6

THE *SPH*<sub>12</sub>~*SST*<sub>13</sub> REGION: NUCLEOTIDE SEQUENCE  
AND LOCATION OF DNA ELEMENTS DIRECTING  
EXPRESSION OF *XYLE* IN pIJ2843 CONSTRUCTS

## 6.1 Introduction

The 965bp *SphI*<sub>12</sub>-*SstI*<sub>13</sub> fragment lies within the 5.5kb *PstI*<sub>8</sub>-*PstI*<sub>18</sub> segment of DNA, which complements the *otcX20* mutation (Figure 6.1) (Butler *et al.*, 1989). The *otcX* class of mutants along with two other classes, *otcY* and *otcZ*, are blocked in the biosynthesis of OTC before the formation of 4-amino-anhydrotetracycline (Rhodes *et al.*, 1981).

The *otcY* mutants are blocked probably in the earliest steps of OTC biosynthesis, as the DNA which complements the *otcY90* mutant has homology with regions of the actinorhodin cluster that encode functions involved in synthesis of the polyketide backbone and its cyclisation to form the isochromanequinone nucleus (see Section 1.14; Butler *et al.*, 1990). Consistent with this proposal, mutants in the class represented by *otcY90* were pale and non-pigmented, while many of the intermediates and shunt products of the pathway are highly coloured (Hunter, pers. comm.).

The position of the *otcZ151* mutation has been mapped close to the "left" of the *PstI*<sub>8</sub> site (Butler *et al.*, 1989). Sequencing of the *BclI*<sub>5</sub>-*SmaI*<sub>9</sub> region (Chapter 4), which includes the *PstI*<sub>8</sub> site, identified a potential protein coding region (PPCR). The predicted product of the *otcZ* gene has high sequence identity to a methyltransferase from bovine pineal glands (Ishida *et al.*, 1987) and a predicted gene product from the region of the tetracenomycin cluster of *S. glaucescens*, which complements *tcmII* and *tcmV* mutants that accumulate C-3 demethyl intermediates (Hutchinson pers. comm.). It is possible, therefore, that the *otcZ151* mutation blocks the methylation of C-6 in the tetracyclic nucleus to form 6-methylpretetramid (6-MPT) (see Section 4.2.3.2).

As the *otcY* and *otcZ* mutants have been assigned provisionally to stages before the formation of 6-MPT, the *otcX20* mutant would appear to be blocked in one of the following; (i) C-4 hydroxylation to form 4-hydroxy-6MPT, (ii) C-6 oxidation to form 4-keto-anhydrotetracycline (4-amino-ATC) or (iii) C-4 transamination to form 4-keto-ATC (see Section 1.14).

It has not yet proven possible to complement the *otcX20* mutation with subclones of this *PstI* fragment (Butler *et al.*, 1990). The 5' end of the *otcZ* structural gene lies at the extreme "left" of the *PstI*<sub>8</sub>-*PstI*<sub>18</sub> region (Figure 6.1). The *otcC* gene, which encodes the anhydrotetracycline (ATC) oxygenase, is located between the *SphI*<sub>12</sub> and *SmaI*<sub>9</sub> restriction sites (Binnie *et al.*, 1989; Section 4.1). As the *otcX20* mutant can produce OTC in cosynthesis tests with the *otcZ151* mutant (Rhodes *et al.*, 1981) and mutations in the *otcC* gene would be blocked in the conversion of anhydrotetracycline to dehydroxytetracycline, the location of the *otcX20* mutation can be delimited to the region between the *SphI*<sub>12</sub> and *PstI*<sub>18</sub> sites (Figure 6.1).

Interestingly, the *SphI*<sub>12</sub>-*SstI*<sub>13</sub> fragment within the 5.5kb *PstI* region has been shown to hybridize at high stringency with a 1.15kb *PstI*-*BglIII* fragment from the *actVa* region of the actinorhodin cluster of *S. coelicolor* A3(2) (Butler *et al.*, 1990; Butler pers. comm.). The *actVa* region (Figure 6.2) is thought to encode functions involved in the hydroxylation of the isochromanone nucleus. The introduction of the *actVa* region into the medermycin producer, *Streptomyces* AM-7161, on low-copy-number vectors (Figure 6.2) led to the production of a hydroxylated form of medermycin, called mederrhodin A (Hopwood *et al.*, 1985b). A mutant in the *actVa* class has been shown also to accumulate kalafungin, which lacks the corresponding hydroxyl group that is present normally on actinorhodin (Cole *et al.*, 1987). Taking all this evidence into account, it could be deduced that the *otcX20* mutant may be blocked in the C-4 hydroxylation of 6-MPT. It is possible, therefore, that the hybridization between the 965bp *SphI*<sub>12</sub>-*SstI*<sub>13</sub> fragment from the *otcX* region and the 1.15kb *PstI*-*BglIII* fragment from the *actVa* region is due to a segment of DNA, which encodes a common function (possibly hydroxylation), which is necessary for the biosynthesis of both oxytetracycline and actinorhodin. The 1.15kb *PstI*-*BglIII* fragment from the *actVa* region has not yet been sequenced completely (Caballero, pers. comm.).



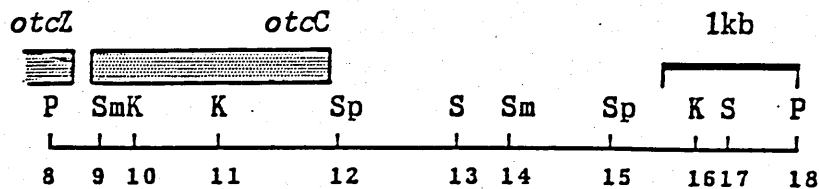


Figure 6.1. The *otcX* region of the oxytetracycline cluster. The hatched boxes (▨) indicate the PCR regions of *otcC* (Butler, unpublished work) and *otcZ* (Chapter 4). Abbreviations: K, *KpnI*; P, *PstI*; S, *SacI*; Sm, *SmaI*; Sp, *SphI*.

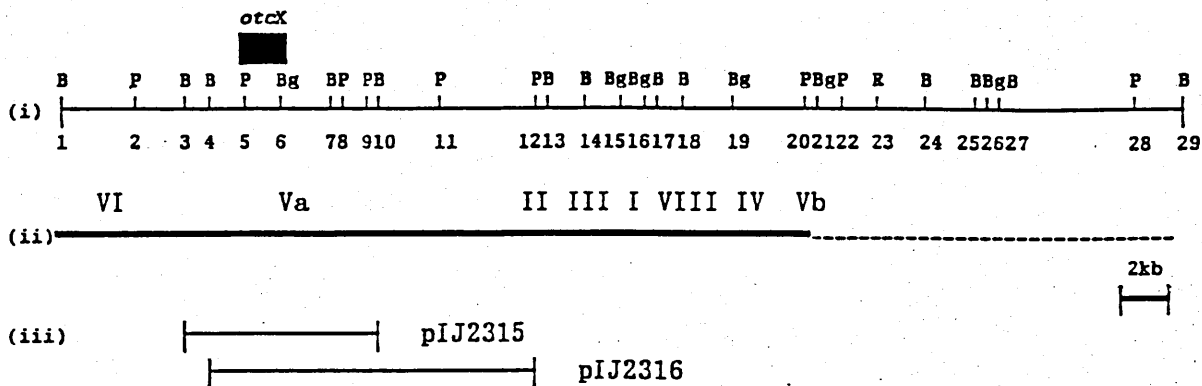


Figure 6.2. Physical and genetic organisation of the *act* cluster in *S. coelicolor* A3(2). The closed box (■) indicates the location of the fragment in the *actVa* region, which hybridizes with the 965bp *SphI*-*SstI* fragment in the *otcX* region of *S. rimosus* (Butler, unpublished work). (i) Restriction map of the *act* region (from Malpartida and Hopwood, 1984). (ii) The positions of the coding regions of the *act* genes and the expected limit of the *act* cluster (bold line, —) were determined by the complementation of mutations using segments of the cluster cloned on plasmid vectors (Malpartida and Hopwood, unpublished result). (iii) The lines (—) show the extent of DNA fragments that resulted in the hydroxylation of medetomidin when introduced into *Streptomyces* AM-7161 on low-copy-number vectors (Hopwood *et al.*, 1985). Abbreviations: B, *BamHI*; Bg, *BglII*; R, *EcoRV*; P, *PstI*. Modified from Hopwood *et al.*, (1986).

The *SphI*<sub>12</sub>-*SstI*<sub>13</sub> fragment was sequenced principally to allow the orientation of the *Sau3AI* inserts, which were cloned in the *xylE*-positive transformants and described in Chapter 5. It was hoped, however, that the comparison of predicted gene products of PPCR(s) in the *SphI*-*SstI* fragment to entries in the NBRF-protein and GenBank databases would identify similarities to polypeptides with known functions, perhaps confirming the role of this region in hydroxylation.

## 6.2 Results

### 6.2.1 DNA Sequencing of the *SphI*-*SstI* Region

The sequencing strategy for the *SphI*<sub>12</sub>-*SstI*<sub>13</sub> region (Figure 6.3) was similar to that used for the *otcZ* gene (Section 4.2.2). The 965bp *SphI*<sub>12</sub>-*SstI*<sub>13</sub> fragment from pPFZ105 was ligated to similarly-digested M13mp18 and M13mp19 (Yanisch-Perron *et al.*, 1985) and introduced into *E. coli* TG1 by genetic transformation (Section 2.2.12). Single-stranded DNA was then isolated from M13 phages, which contained recombinant DNA molecules (Section 2.3.16). Taq polymerase, as supplied in the TAQuenase™ kit by USB (Section 2.3.17), was used to sequence the single-stranded templates mKM810 and mKM910, which were derived from the insertion of the *SphI*-*SstI* fragment into M13mp18 and M13mp19 respectively. Nucleotide mixes containing 7-deaza-dGTP (Mizusawa *et al.*, 1986) were used to minimise the formation of secondary structure in the extension products, which would lead to band compressions in the sequencing ladders. As the standard sequencing ladders could only be read for around 350nt, oligonucleotides were synthesized and used as primers to extend the sequence.

The *SphI*<sub>12</sub> boundary was sequenced from both strands using the templates, mL6A and mL6C, and oligonucleotides, C64 and X84, which were used to map the 5' termini of the divergent transcripts (Section 5.2.4.1). Sequencing of the *Sau3AI* insert from the pIJ2843 construct in isolate 15 (templates mKM8X15 and mKM9X15) spanned the *SstI*<sub>13</sub> site.

In total, 1166bp was sequenced between the annealing site of oligo. C64 (114bp to the left of the *SphI*<sub>12</sub> site) and the *Sau3AI*<sub>g</sub> site (87bp to the right of the *SstI*<sub>13</sub> site). Over 85% of the DNA was sequenced on both strands. The sequence was determined unambiguously in the regions, which were sequenced from only one strand.

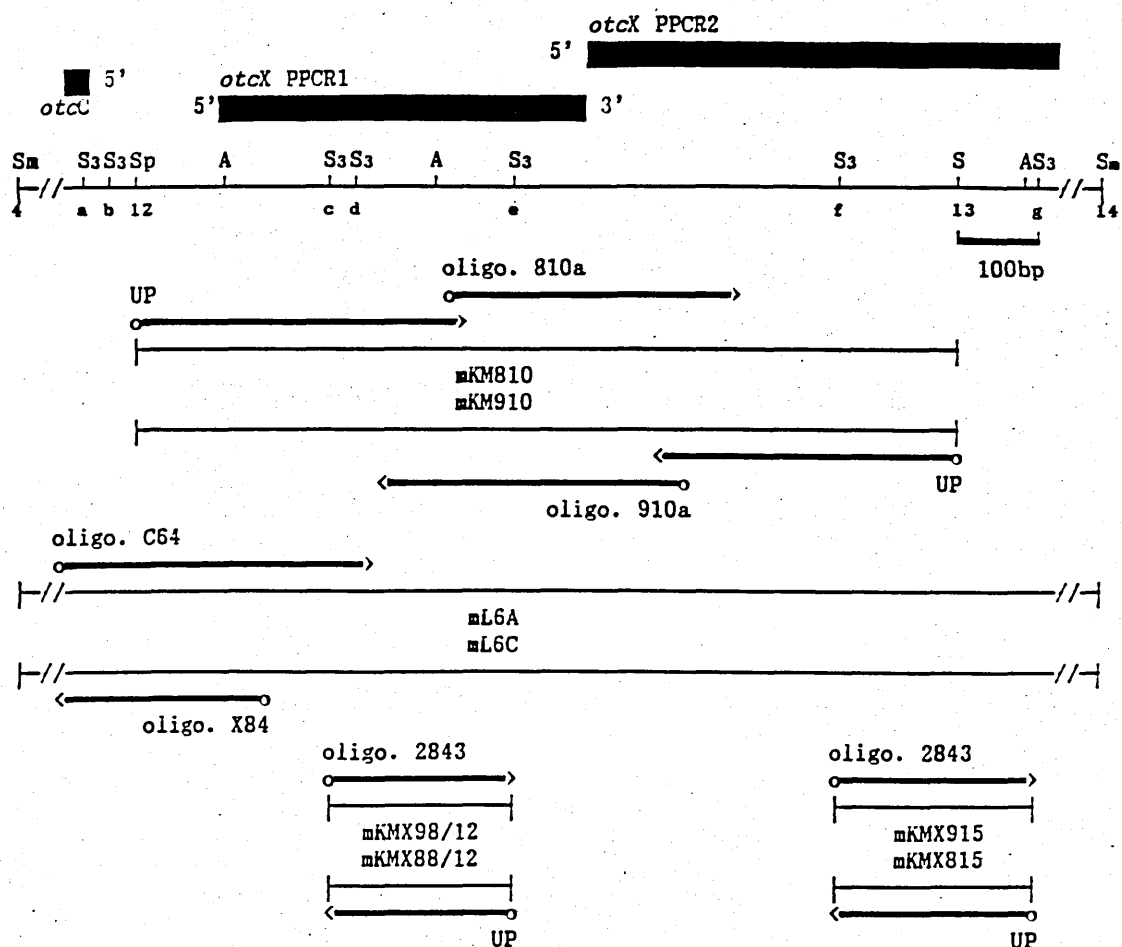


Figure 6.3. Strategy used to sequence the 965bp *Sph*I<sub>12</sub>-*Sst*I<sub>13</sub> Region. The lines (|---|) indicates the region of DNA cloned in the M13 constructs. The arrows (o==>) show the direction and extent of the nucleotide sequence analysis. The oligonucleotides used to prime the extension reactions are given above the arrows. Abbreviations: UP, universal (-20) primer ; oligo., oligonucleotide.

The average mol G+C composition of this region is 72.0%. The annotated sequence, including the predicted amino acid sequences of potential protein coding regions, potential ribosome binding sites, the oligonucleotides used to prime the sequencing reactions and the restriction sites referred to in the text are shown in Figure 6.4.

### 6.2.2 Features of the Nucleotide Sequence

The 1166bp region containing the *SphI*-*SstI* fragment was analysed primarily using the "Sequence Analysis Software Package" (Version 6.2) from the Genetics Computer Group at the University of Wisconsin (UWGCG). The individual UWGCG programs have been described already (Section 4.2.3) and thus, will not be discussed in detail within this section.

The sequence was analysed for potential protein coding regions (PPCRs) using the program FRAME (Bibb *et al.*, 1984). A non-random distribution of G+C nucleotides was detected in three different segments of the 1166nt sequence (Figure 6.5). Analysis of the sequence for unequal usage of synonymous codons using the UWGCG program TESTCODE also indicated that the same regions encoded proteins (result not shown). Analysing the sequence from the annealing site of oligo. C64, the FRAME plots were diverse immediately. By nucleotide position 100, the plots had converged to around 68% mol G+C, suggesting that a PPCR was present at the extreme left of the sequence. This PPCR corresponds to the 5' end of the *otcC* gene (Section 4.1) (Binnie *et al.*, 1989). A potential ribosome binding site (GAGGAG) with a high degree of complementarity to the 3' end of the 16S rRNA from *S. lividans* (sequenced by Bibb and Cohen, 1982) was located 5nt upstream of the AUG start codon (69-67nt) for *otcC*, which is in nucleotide frame N2< (Figure 6.5). The predicted amino acid sequence from the 5' end of *otcC* is in complete agreement with the sequence obtained from the amino terminus of the gene product, ATC oxygenase (Binnie *et al.*, 1989).

From base position 100, the FRAME plots diverge once more. The thin line plot, which commences from base 1 fluctuates around 70% and thus, corresponds to the first position of the codons in this PPCR.

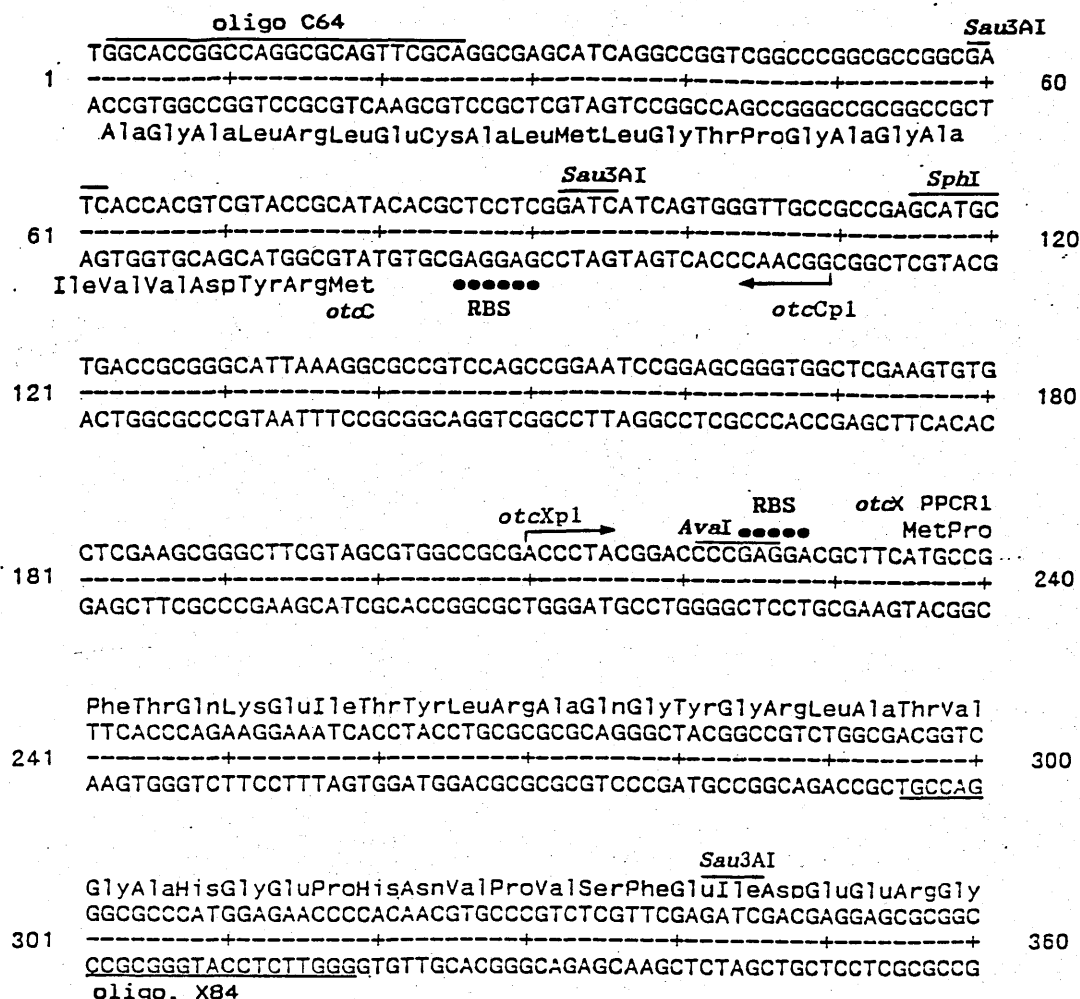


Figure 6.4. Annotated sequence of the 965bp *SphI*<sub>12</sub>-*SstI*<sub>13</sub> region. The predicted amino acid sequence of *otcX* PCR1 and PCR2 are shown above the corresponding nucleotide sequences. Also indicated are the nucleotides incorporated into the primers used to further the sequence, possible ribosome binding sites (RBS) for *otcC*, *otcX* PCR1 and PCR2 and the transcription start points for *otcCp1* and *otcXp1*. For clarity, only restriction sites referred to in the text are indicated. Continued on pages 181 and 182.

Sau3AI

ThrIleGluIleThrGlyArgAspMetGlyArgSerArgLysPheArgAsnValAlaLys  
 361 ACCATCGAGATCACCGGCCGGGACATGGGCCGCGAGCCGCAAGTTCGGGAACGTGGCGAAG 420  
 TGGTAGCTCTAGTGGCCGGCCCTGTACCCGGCGTCGGCGTTCAAGGCCTTGCACCGCTTC

AvaI

AsnAspArgValAlaPheValValAspAspValProCysArgAspProGluValValArg  
 421 AACGACCGGGTCGCCTTCGTCTGGACGACGTGCCGTGCCGCGACCCCGAGGTGGTCCGC 480  
 TTGCTGGCCCAGCGGAAGCAGCACCTGCTGCACGGCACGGCGCTGGGGCTCCACCAGGCG

AlaValValIleHisGlyThrAlaGlnAlaLeuProThrGlyGlyArgGluArgArgPro  
 481 GCCGTCGTCATCCACGGCACGGCCAGGCGCTGCCACCGGGCGGCAGGGAGCGCCGGCCG 540  
 CGGCAGCAGTAGGTGCCGTGCCGGGTCCGCGACGGGTGGCCGCCGTCCCTCGCGGCCGGC

Sau3AI

HisCysAlaAspGluMetIleArgIleHisProArgArgIleValThrTrpGlyIleGlu  
 541 CACTGCGCCGACGAGATGATCCGCATCCACCCCGGGCGCATCGTCACTGGGGCATCGAA 600  
 GTGACGCGGCTGCTCTACTAGGCGTAGGTGGGGGCCGCGTAGCAGTGGACCCCGTAGCTT

RBS

GlyAspLeuSerThrGlyValHisAlaArgAspIleThrAlaGluAspGlyGlyArgArg  
 601 GGGGACCTGTCCACCGGGGTCCACGCACGCGACATCACGGCCGAGGACGGAGGACGGCGA 660  
 CCCCTGGACAGGTGGCCCCAGGTGCGTGCGCTGTAGTGCCGGCTCCTGCCTCCTGCCGCT

*otcX* PPCR2

etArgGluLeuProGluPheProTrpAspValLeuLeuProTyrLysLysArgAlaAlaA  
 End  
 661 TGAGGGAACTGCCCGAATTTCCGTGGGATGTGCTGCTGCCGTACAAGAAGCGCGCCGCGG 720  
 ACTCCCTTGACGGGCTTAAAGGCACCCTACACGACGACGGCATGTTCTTCGCGGGCGGCC

IaHisProAspGlyLeuValAsnLeuAlaLeuGlyGluProValGluAlaThrProAspV  
 721 CCCACCCCGACGGGCTGGTCAACCTGGCGCTGGGCGAACCAGGTGCAAGCCACGCCGACG 780  
 GGGTGGGGCTGCCCCACCAGTTGGACCGCGACCCGCTTGGCCAGCTTCGGTGCGGGCTGC

840

900

960

1020

1080

1140

100

The beaded line (—●—) originating from base 2 reaches an average G+C composition of 50% by base position 200 and thus, corresponds to the second codon position. Similarly, the heavy line from base 3, which fluctuates around 95%, corresponds to codon position 3. Thus, the open reading frame for *otcX* PPCR1 is translated from left to right in nucleotide frame 1 (N1>), which is divergent to *otcC*.

The AUG codon at base position 235 is probably the translation start codon for *otcX* PPCR1, as it is within the region of plot divergence and it is preceded by a potential ribosome binding site (GAGGA), which is located 6nt upstream. PPCR1 ends at base position 663 with a UGA codon. As expected for a gene from *Streptomyces*, *otcX* PPCR1 has an average mol G+C distribution in the first, second and third codon positions of 69.9%, 49.0% and 92.3% respectively. In accordance with the assignment of PPCR1 to the region between base positions 235 and 663, the FRAME plots start to converge around base position 500. The FRAME plots do not fluctuate, however, around 70% mol G+C composition, instead they diverge once more. The thin line plot, which commences from base 1, reaches a position where it fluctuates around 50% mol G+C composition, suggesting that it corresponds to the second codon position of another PPCR. The first codon position was assigned to the heavy line plot originating from base 3, as it fluctuates around 75%. Similarly the beaded line (—●—) from base 2 fluctuates around 90% mol G+C composition and was assigned to the third codon position. The open reading frame for *otcX* PPCR2 is translated, therefore, from left to right in nucleotide frame N>3.

Within the region of plot divergence (base position 600-700) an AUG codon (660-662) and a GUG codon (690-692) were located within N>3. The translation start codon of *otcX* PPCR2 was assigned to the AUG codon as it was preceded (6nt upstream) by a reasonable RBS (GGAGG). The GUG codon was not preceded by a sequence with significant complementarity to the 3' end of 16S rRNA from *S. lividans* (Bibb and Cohen, 1982). PPCR2 does not contain an inframe stop codon. This incomplete PPCR has a typical average mol G+C composition in the first, second and third codon positions of 77%, 54.2% and 90.5% respectively.



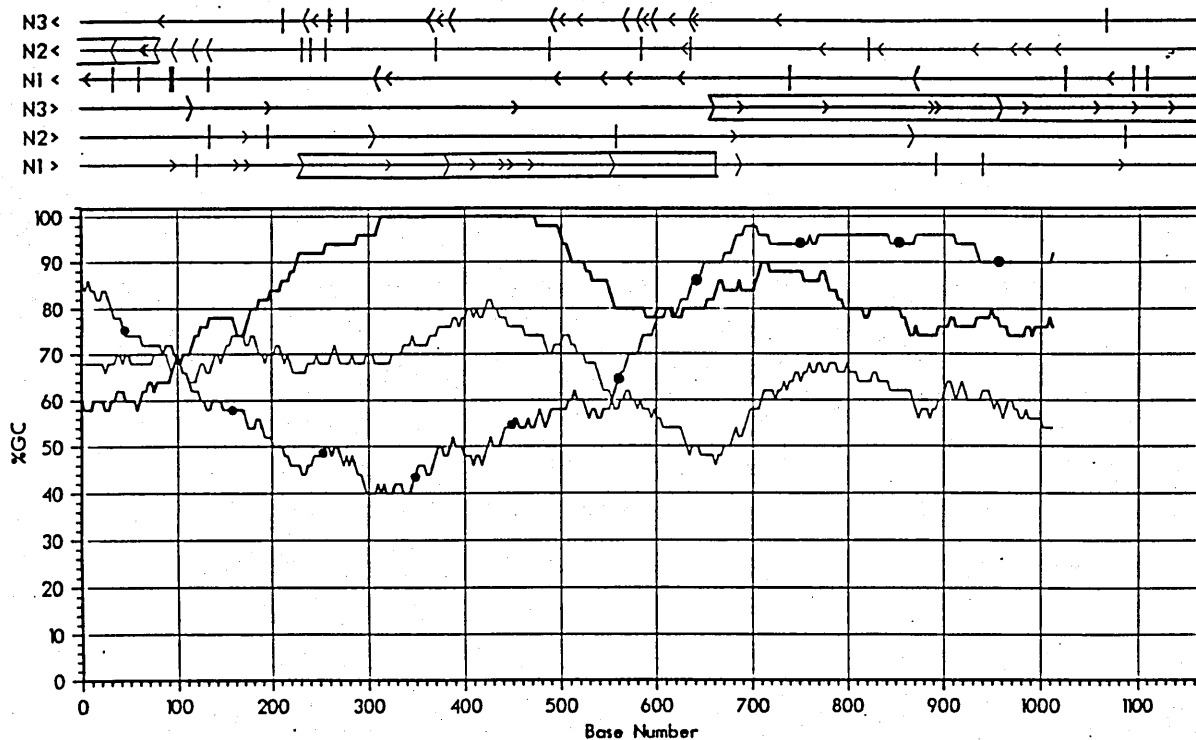


Figure 6.5. FRAME plot (Bibb *et al.*, 1984) of the 965bp *Sph*I<sub>12</sub>-*Sst*I<sub>13</sub> region using a "window" of 50 triplets. The thin line plot (—) represents the first triplet position and its associated reading frames (N1> and N2<); the beaded line (—●—) represents the second triplet position and its associated reading frames (N2> and N3<) and the heavy line (—) indicates the third triplet position and its associated reading frames (N3> and N1<). Symbols > and < indicate AUG codons in the forward and reverse frames respectively. Symbols > and < indicate GUG codons and the symbol | indicates stop codons.

The putative AUG start codon (660-662) for PPCR2 overlaps with the UGA stop codon (661-663) for PPCR1. It is possible, therefore, that PPCR1 and PPCR2 are translationally coupled similar to those found in a variety of prokaryotic operons (Zalkin and Ebbole, 1988; Harms *et al.*, 1988). Such an arrangement would facilitate the synthesis of proteins encoded by PPCR1 and PPCR2 in stoichiometric amounts. The products of these PPCRs could be components of a multi-peptide complex. The translation initiation of *otcX* PPCR1 and PPCR2 is considered further in Chapter 7.

### 6.2.3 Analysis of the Deduced Amino Acid Sequence of PPCR1 and PPCR2

The predicted gene product of PPCR1 is composed of 142 amino acids and has a calculated  $M_r$  of 15,851. The incomplete PPCR2 encodes 169 amino acids. Comparison of the codon usage in PPCR1 and PPCR2 with a codon preference table compiled from 63 streptomycete genes (Bibb, unpublished data) revealed that the biased usage of synonymous codons within the PPCRs of *otcX* was in accordance with the codon preferences observed for other *Streptomyces* genes (Figure 6.5).

In an attempt to determine the possible functions of the polypeptides encoded in the 1166bp region containing the *SphI*<sub>12</sub>-*SstI*<sub>13</sub> fragment, the predicted amino acid sequences were compared with the entries in the NBRF-protein database (Release 25, June 1990) and the GenBank database (Release 60, June 1989) using the UWGCG programs WORDSEARCH/SEGMENTS and TFASTA. No striking similarities were found.

### 6.2.4 Location of the DNA elements from the *SphI*-*SstI* region which direct transcription of *xylE* in pIJ2843 Constructs

In the previous chapter, the *PstI*<sub>8-18</sub> region was digested partially with *Sau3AI*, ligated with the promoter-probe vector pIJ2843, which had been digested with *BamHI*, and then introduced into *S. lividans* TK54 by genetic transformation. Several transformants contained recombinant vectors with inserts from the *SphI*-*SstI* region, which were capable of directing transcription of the *xylE* reporter gene.

(A) BIBB TABLE

	T	C	A	G	
T	F TTT 0.04	S TCT 0.06	Y TAT 0.11	C TGT 0.10	T
	F TTC 2.89	S TCC 2.11	Y TAC 2.22	C TGC 0.69	C
	L TTA 0.02	S TCA 0.15	* TAA 0.01	* TGA 0.26	A
	L TTG 0.26	S TCG 1.40	* TAG 0.05	W TGG 1.55	G
C	L CTT 0.22	P CCT 0.18	H CAT 0.16	R CGT 0.51	T
	L CTC 3.52	P CCC 2.36	H CAC 2.33	R CGC 3.74	C
	L CTA 0.03	P CCA 0.07	Q CAA 0.18	R CGA 0.32	A
	L CTG 5.08	P CCG 2.87	Q CAG 2.45	R CGG 2.96	G
G	I ATT 0.14	T ACT 0.14	N AAT 0.10	S AGT 0.14	T
	I ATC 3.07	T ACC 4.49	N AAC 2.35	S AGC 1.46	C
	I ATA 0.10	T ACA 0.15	K AAA 0.12	R AGA 0.10	A
	M ATG 1.55	T ACG 1.99	K AAG 2.29	R AGG 0.46	G
A	V GTT 0.27	A GCT 0.38	D GAT 0.27	G GGT 0.74	T
	V GTC 4.40	A GCC 7.77	D GAC 6.16	G GGC 0.74	C
	V GTA 0.16	A GCA 0.55	E GAA 1.11	G GGA 0.85	A
	V GTG 3.08	A GCG 4.39	E GAG 4.83	G GGG 1.70	G

(B) *otcX* PPCR1

	T	C	A	G	
T	F TTT 0.00	S TCT 0.00	Y TAT 0.00	C TGT 0.00	T
	F TTC 2.80	S TCC 0.70	Y TAC 1.40	C TGC 1.40	C
	L TTA 0.00	S TCA 0.00	* TAA 0.00	* TGA 0.70	A
	L TTG 0.00	S TCG 0.70	* TAG 0.00	W TGG 0.70	G
C	L CTT 0.00	P CCT 0.00	H CAT 0.70	R CGT 0.70	T
	L CTC 0.00	P CCC 3.50	H CAC 3.50	R CGC 6.99	C
	L CTA 0.00	P CCA 0.00	Q CAA 0.00	R CGA 0.70	A
	L CTG 2.80	P CCG 2.10	Q CAG 2.10	R CGG 4.20	G
G	I ATT 0.00	T ACT 0.00	N AAT 0.00	S AGT 0.00	T
	I ATC 0.70	T ACC 4.90	N AAC 2.10	S AGC 0.70	C
	I ATA 0.00	T ACA 0.00	K AAA 0.00	R AGA 0.00	A
	M ATG 2.10	T ACG 2.10	K AAG 2.10	R AGG 0.70	G
A	V GTT 0.00	A GCT 0.00	D GAT 0.00	G GGT 0.00	T
	V GTC 6.29	A GCC 4.20	D GAC 6.99	G GGC 6.99	C
	V GTA 0.00	A GCA 0.70	E GAA 2.10	G GGA 2.10	A
	V GTG 3.50	A GCG 2.80	E GAG 5.59	G GGG 1.40	G

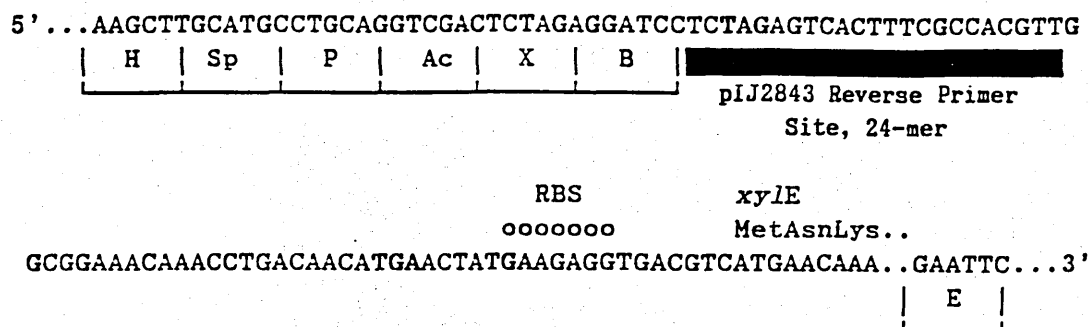
(C) *otcX* PPCR2

	T	C	A	G	
T	F TTT 1.79	S TCT 0.00	Y TAT 0.00	C TGT 0.00	T
	F TTC 0.60	S TCC 1.19	Y TAC 1.19	C TGC 0.00	C
	L TTA 0.00	S TCA 0.60	* TAA 0.00	* TGA 0.00	A
	L TTG 0.00	S TCG 1.19	* TAG 0.00	W TGG 2.38	G
C	L CTT 0.00	P CCT 0.00	H CAT 0.00	R CGT 0.00	T
	L CTC 3.57	P CCC 7.14	H CAC 0.60	R CGC 3.57	C
	L CTA 0.00	P CCA 0.00	Q CAA 0.00	R CGA 0.00	A
	L CTG 9.52	P CCG 5.95	Q CAG 0.00	R CGG 1.79	G
G	I ATT 0.00	T ACT 0.00	N AAT 0.00	S AGT 0.60	T
	I ATC 1.19	T ACC 2.38	N AAC 1.79	S AGC 0.60	C
	I ATA 0.00	T ACA 0.00	K AAA 0.00	R AGA 0.00	A
	M ATG 1.19	T ACG 2.98	K AAG 2.38	R AGG 0.60	G
A	V GTT 0.00	A GCT 0.00	D GAT 1.19	G GGT 0.00	T
	V GTC 3.57	A GCC 5.36	D GAC 4.17	G GGC 6.55	C
	V GTA 1.19	A GCA 0.60	E GAA 3.57	G GGA 0.00	A
	V GTG 4.76	A GCG 7.74	E GAG 3.57	G GGG 1.79	G

Figure 6.6. (A) Codon usage in a file of 63 streptomycete genes (compiled by Bibb, unpublished data). (B) Codon usage in *otcX* PPCR1. (C) Codon usage in the incomplete PPCR2. The numbers represent the frequency with which the codons are used. The frequency is expressed as a percentage (%) of the total number of codons in the file. Amino acids are indicated by the standard single letter code and stop codons by asterisks.

To delineate the locations of the promoter elements within the *SphI*-*SstI* region, the *Sau3AI* fragments from recombinant vectors in the *xylE*-positive transformants 8, 12 and 15 (see Table 5.1) were sequenced. The *Sau3AI* inserts were removed on *HindIII*-*EcoRI* fragments and ligated with similarly-digested M13mp18 and M13mp19. The *HindIII*-*EcoRI* fragments also contain the *xylE* reporter gene, which is located between the former *BamHI* site (used to insert the *Sau3AI* fragments) and the *EcoRI* site. The ligation mixes were then introduced into *E. coli* TG1 by genetic transformation (Section 2.2.12). Single-stranded templates were isolated from M13 phage containing recombinant DNA molecules (Section 2.3.16). The *HindIII*-*EcoRI* fragment from the pIJ2843 construct in isolate 8 was used to derive mKMx88 and mKMx98 from M13mp18 and M13mp19 respectively. Similarly mKMx812/912 and mKMx815/915 were derived from the *HindIII*-*EcoRI* fragments of isolates 12 and 15 respectively. The *Sau3AI* inserts in the M13mp19 constructs could not be sequenced from the universal (-20) primer, as the *xylE* reporter gene was located between the primer annealing site and the insert. These constructs were sequenced, therefore, from an oligonucleotide, which anneals to a complementary sequence on the antisense strand of *xylE*, immediately downstream of the multiple cloning site in pIJ2843 (Figure 6.7).

Sequencing revealed that the recombinant vectors in isolates 8 and 12 both contained the *Sau3AI*<sub>c-e</sub> fragment with the *Sau3AI*<sub>c</sub> site nearest the *xylE* reporter gene. Should the DNA element, which is directing the transcription of *xylE* in the pIJ2843 constructs in isolates 8 and 12, be active in the chromosome of *S. rimosus* it would direct transcription towards the *SphI*<sub>12</sub> site. In other words, it would produce transcripts from the antisense strand of *otcX* PPCR1. Similarly, the pIJ2843 construct in isolate 15 contains the *Sau3AI*<sub>f-g</sub> fragment, with the *Sau3AI*<sub>f</sub> site nearest the *xylE* reporter gene. Should the promoter element in this insert be active in the chromosome it would direct the synthesis of transcripts from the antisense strand of *otcX* PPCR2.



**Figure 6.7.** Multiple Cloning Site in pIJ2843. The DNA strand, which is shown, is generated when the *Hind*III-*Eco*RI fragment is inserted into M13mp19 and isolated as single-stranded DNA. The nucleotides in bold type are the three translation stops codons. Abbreviations: B, *Bam*HI; Ac, *Acc*I; E, *Eco*RI; H, *Hind*III; P, *Pst*I; S, *Sac*I; Sp, *Sph*I; X, *Xba*I.

### 6.2.5 Analysis of transcription within the *Sau3A*<sub>Ic-e</sub> and *Sau3A*<sub>Ic-g</sub> regions

Single-stranded probe, which was complementary to the sense strand of *otcX* PPCR1 within the *Sau3A*<sub>Ic-e</sub> region, was generated by annealing the pIJ2843 Reverse Primer (Figure 6.7) to mKMX912 (Figure 6.3), extending from the oligonucleotide in the presence of radio-labelled nucleotides and digesting with *Hind*III (Materials and Methods, 2.3.18). Similarly, a probe complementary to sense strand of *otcX* PPCR2 within the *Sau3A*<sub>Ic-g</sub> region was generated from mKMX915. These continuously-labelled probes were hybridized initially at 37°C with total RNA from *S. rimosus* M4018 grown on TSB, *S. rimosus* M15883 grown on LCM, which produces low titres of OTC and *S. rimosus* M15883 grown on the high-titre production medium, TS1. This selection of RNA was the same as that described in Section 3.2.4. After digestion with S1 nuclease, no RNA-protected fragments were detected in any of the samples. This procedure was repeated, carrying out the hybridizations at 22°C. Although transcripts overlapping with the probes by as little as 15-20nt should have formed stable hybrids, no RNA-protected fragments were detected after digestion with S1 nuclease.

### 6.3 Discussion

It has been proposed that overlapping bidirectional protein coding regions are present in *Streptomyces*. For example, ORF85 and ORF79 in the plasmid pIJ101 (Kendall and Cohen, 1988) and ORFs 1215 and 330 in the Insertion Element IS110 (Bruton and Chater, 1987). In both of these cases, however, there is a compromise in the codon usage of the protein coding regions in the area of overlap. The average mol G+C composition in the first, second and third codon positions are approximately 70%, 45%, and 70% respectively. The first codon position of the forward PPCR is the third codon position of the reverse PPCR and *vice versa*. The average distribution of G+C residues in the codons of *otcX* PPCR1 and PPCR2 is typical of *Streptomyces* genes. As the typical codon usage in PPCR1 and PPCR2 would necessitate the use of rare codons in the antisense RNA it was considered unlikely that transcripts originating from the element(s) cloned in isolates 8, 12 and 15 would be translated efficiently, unless the antisense transcripts encoded very

short polypeptides (ca. 50 amino acids). The FRAME analysis would not detect necessarily a compromise in the distribution of G+C residues in the different codon positions within a limited segment of overlapping PPCRs. An analysis of the 1166bp region did not identify, however, any reverse ORFs (reading right to left) containing inframe translation start codons at the 5' end, which were preceded by consecutive nucleotides with complementarity to the 3' end of 16S rRNA from *S. lividans* (Bibb and Cohen, 1982). As no evidence has been obtained for the presence of protein coding regions in the reverse frames of the *SphI*-*SstI* region, it seemed unlikely that antisense transcripts arising from the promoter elements cloned in isolates 8, 12 and 15 would be messenger RNAs.

Increasingly, small RNA molecules are being implicated in regulating translation, as well as participating in the control of transcription, transposition, phage development, DNA replication, plasmid copy number and plasmid transfer (for review and examples, see Inouye and Delihias, 1988). Often an antisense RNA transcript, which is complementary to the translation start codon and Shine-Delgarno (SD) sequence, binds the mRNA preventing the initiation of translation, for example: pOUT RNA binds to the SD sequence and AUG start codon of transposase mRNA and thereby, regulates Tn 10 transposition (Simons and Kleckner, 1983); conjugal transfer of plasmid R100 is regulated by two *finP* transcripts which are envisaged to block translation of the *traJ* message by binding to the SD region (Dempsey, 1987) and the translation of the antiterminator (*ant*) mRNA in bacteriophage P22 is inhibited by the binding of a small RNA, which shields the AUG codon and SD sequence (Liao *et al.*, 1987). Not all antisense RNA transcripts regulate transcription by preventing ribosome binding, for example, in bacteriophage lambda OOP RNA prevents *cII* expression by hybridizing to the 3' end of the message and creating an RNAase III-sensitive site (Krinke and Wulff, 1987). As the putative antisense transcripts in the *otcX* region are unlikely to encode polypeptides, they could conceivably regulate the translation of a *otcX* message, which originates from *otcX*p1 (Figure 6.4), by an unknown mechanism.

The identification of conserved *pho*-like boxes in the promoters *otcCp1* and *otcXp1* (Section 5.2.7), however, has already lead to the proposal of a plausible model for regulating the expression of genes in the *otc* cluster. Transcription from *otcCp1* and *otcXp1* and possibly other promoters within the *otc* cluster, which direct transcription of genes whose products are involved directly in the biosynthesis of OTC, may be activated by the binding of a regulatory protein within the promoter region. Although similar models involving *trans*-acting regulatory proteins have been proposed to regulate the biosynthesis of other antibiotics (for review, see Section 1.6), to date, no articles have been published which suggest a role for antisense RNA in regulating the expression of enzymes of secondary metabolism at the level of translation in *Streptomyces*.

In *Bacillus subtilis*, the genes encoding enzymes which have roles in the early steps of arginine biosynthesis are located in a single cluster, with the gene order *argC-argA-argE-argB-argD-cpa-argF* (Mountain, et al., 1986). Transcriptional analysis *in vitro* revealed a promoter upstream of *argC*, which directed transcription rightwards. The insertion of small DNA fragments into the coding region of *argC* had a polar effect, which disrupted the expression of *argF* (Smith, unpublished result). As this *arg* cluster is not thought to contain a regulatory locus, it was deduced that *argC-argA-argE-argB-argD-cpa* and *argF* were part of a single transcriptional unit (Baumberg and Smith, pers. comm.) However, the shotgun cloning of random fragments from a segment of DNA, which contained the proposed *argC-argA-argE-argB-argD-cpa-argF* operon, into a promoter-probe vector identified four internal fragments that were capable of directing transcription of the reporter gene (Baumberg, unpublished result). As there are no precedents for the antisense RNA regulation of gene expression in *B. subtilis*, it was proposed that the cloned elements may be "phantom" promoters, which have no role in the transcription or regulation of the *arg* operon in the chromosome (Baumberg, pers. comm.). By analogy with the *arg* operon, it is a formal possibility that the plasmid-active promoter (PAP) elements cloned in isolates 8, 12 and 15 are not active in the chromosome of *S. rimosus*. The inability to detect the 5' ends of transcripts originating from the *Sau3A*I<sub>c-e</sub> and *Sau3A*I<sub>f-g</sub> regions in total RNA isolated from *S. rimosus* at different stages in the fermentation



of OTC is consistent with the PAP elements in these regions being "phantom" promoters. As will be discussed later, there are other possible reasons why the 5' ends of the putative antisense transcripts were not detected. At this stage in the analysis it is not clear what roles, if any, the promoter elements, cloned in isolates 8, 12 and 15 have in the biosynthesis of OTC.

The 259bp *Sau3A*I<sub>b-c</sub> fragment containing the *otcCp1/otcXp1* promoter region (see Figure 6.4) has been inserted into the *Bam*HI site of pIJ2843 to derive pKMSOCp1, which has the *Sau3A*I<sub>b</sub> site nearest the *xylE* reporter gene (McDowall and Kok, unpublished work). Transformants of *S. lividans* TK54, which contain pKMSOCp1, only express *xylE* after incubation for approximately six days on Emersons agar (containing thiostrepton, 25µg.ml<sup>-1</sup>) at 30°C (McDowall and McGregor, unpublished result). A similar delay in transcription from the *otcCp1* promoter is also observed when the same transformants were incubated on R2 medium (McDowall and Kok, unpublished result). In the previous chapter, the assays for transformants containing recombinant pIJ2843 vectors with promoter elements capable of directing transcription of the *xylE* reporter gene were carried out after incubation for 3, 4 and 5 days on regeneration plates at 30°C (Section 5.2.2). At each time point, six *xylE*<sup>+</sup> isolates were selected for further analysis. This screening strategy would be biased, therefore, against promoters which were only active after incubation for 6 days, given that all the isolates which were selected (with the exception of isolate 5) were highly pigmented after spraying with catechol at each of the time points (Section 5.2.3). As the *otcXp1* and *otcCp1* promoters have been shown to have similar temporal activities in *S. rimosus* (Section 5.2.6), it is perhaps not surprising that isolates containing pIJ2843 constructs with the *Sau3A*I<sub>b-c</sub> fragment in either orientation were not selected.

The S1 mapping experiments described in Section 6.2.5 failed to detect transcripts originating from either the *Sau3A*I<sub>c-e</sub> or *Sau3A*I<sub>f-g</sub> regions in total RNA isolated from *S. rimosus* at different stages in the fermentation of oxytetracycline. The possibility that the elements cloned in isolates 8, 12 and 15 were "phantom" promoters has already been discussed. It is also possible that the 5' ends of the putative antisense transcripts were not detected, as they did not overlap to any

great extent with the single-stranded probes (Section 6.2.5). Interestingly, the *Sau3A*<sub>10-e</sub> fragment cloned in isolates 8 and 12 was the product of partial digestion. It is possible that the internal *Sau3A*<sub>10</sub> site, which is located only 31bp to the right of the *Sau3A*<sub>10</sub> site, lies within sequences required to direct transcription of *xylE* in the pIJ2843 constructs. Should this be the case, then the transcription start point of the promoter element may well be located close to the *Sau3A*<sub>10</sub> site and thus transcripts originating from this element might not have hybridized efficiently to the probe generated from mKMx912 (Figure 6.3).

Primer extension from oligonucleotide C64 hybridized to total RNA isolated from *S. rimosus* (Section 5.2.5) should have detected the 5' ends of transcripts originating within the *Sau3A*<sub>10-e</sub> region, provided transcription did not terminate before the primer annealing site. Similar to the rho factor-independent transcription termination sites in the enterobacteria (for review, see Galloway and Platt, 1986), the putative transcription termination sites of many streptomycete genes contain G+C-rich dyad symmetry (for review and references, see Seno and Baltz, 1988). Streptomycete terminators do not contain, however, the string of uridine residues, which follows the region of dyad symmetry at the 3' end of transcripts in the enterobacteria. No sequences with G+C-rich dyad symmetry typical of streptomycete terminators were detected in the intervening region between the *Sau3A*<sub>10</sub> site and the annealing site for oligo. C64 (Figure 6.4). In the enterobacteria, another class of termination sites, which do not possess dyad symmetry and a string of uridine residues, can also cause RNA polymerase to terminate transcription. Although these termination sites are generally A+T-rich and require additional protein factors, there is no unifying sequence or structure for this second class of terminators (for reviews, see Platt and Bear, 1983). As yet, no factor-dependent termination sites have been characterised in *Streptomyces*. It remains a formal possibility, therefore, that transcripts originating from the element(s) cloned in isolates 8 and 12 terminate at an unidentified factor-dependent terminator before the annealing site for oligo. C64. Another explanation for why the 5' ends of the putative antisense transcripts were not detected using primer extension analysis, is that the cloned promoter elements might only be active transiently. This might be expected if the antisense transcripts have a role in regulating the expression of the

*otcX* locus and thus, the biosynthesis of OTC. No attempt, as yet, has been made to correlate the presence of catechol dioxygenase activity in plate assays (Section 5.2.3) with the intracellular levels of *xylE* transcripts, originating from cloned promoters in pIJ2843 constructs. The detection of catechol dioxygenase activity in isolates 8, 12 and 15 after incubation for 3, 4 and 5 days (Chapter 5, 5.2.3) is not inconsistent necessarily, therefore, with the promoter elements being active transiently.

On balance, the analyses described above suggest that the elements cloned in transformants 8, 12 and 15 are unlikely to direct the synthesis of mRNAs, which encode products involved directly in the biosynthesis of OTC. Although the above discussion does not lead to any firm conclusion about the role of these promoter elements, it highlights the range of possibilities which must be considered in undertaking further transcriptional analysis of the *otcX* locus. The lines of investigation which are being considered to characterize further the promoter elements cloned in isolates 8, 12 and 15 are described in the concluding chapter.

It was hoped that the comparison of the amino acid sequences of the predicted products from *otcX* PPCR1 and PPCR2 with entries in the NBRF-protein and EMBL databases would identify similarity to polypeptides with known functions and thus suggest a possible role for the *otcX* locus in the biosynthesis of OTC. Unfortunately, no similarities to any entries in the databases were detected.

The *actVa* region in the actinorhodin cluster of *S. coelicolor* is currently being sequenced (Caballero, pers comm.). Comparison of sequence in the 0.96kb *SphI*<sub>12</sub>-*SstI*<sub>13</sub> fragment from the *otcX* region (Figure 6.1) and the 1.15kb *PstI*<sub>5</sub>-*BglII*<sub>6</sub> fragment from the *actVa* region (Figure 6.2) should reveal the DNA determinants which permit hybridization between these regions at high stringency.

## **CHAPTER 7**

### **TRANSLATION INITIATION**

## 7.1 Introduction

*E. coli* has been used as a model for studying translation in prokaryotic systems. Translation initiation is specified by sequences in messenger RNA other than start codons. The initiation codons of many genes from *E. coli* are preceded by a short sequence of nucleotides that are directly complementary to some part of the -ACCUCCUUA-OH nucleotide sequence, lying at the extreme 3' end of the 16S ribosomal RNA of *E. coli* (Shine and Dalgarno; 1974, 1975).

The 16S rRNA is an integral component of the 30S ribosomal subunit, which interacts with mRNA in conjunction with ribosomally associated proteins and initiation factors. Shine and Dalgarno postulated that a preliminary interaction between sequences at the 3' end of 16S rRNA and complementary sequences on the mRNA could target a 30S ribosomal subunit towards an AUG or GUG translation initiation codon. This interaction would allow the formation of a 30S subunit-mRNA-f<sup>met</sup>tRNA complex, which was positioned correctly and could give rise to an active in-frame ribosome on addition of the 50S ribosomal subunit. The translation of mRNA can be affected by the complementarity of the Shine-Dalgarno (SD) sequence and the spacing of this sequence from initiation codons (for review, see Gold and Stormo, 1987).

For the Gram-positive genera, *Bacillus* and *Staphylococcus*, it has been proposed that a high degree of complementarity between the sequence at the 3' end of 16S rRNA and sequence immediately upstream of the translation initiation codon on the mRNA (equivalent to the SD sequence in *E. coli*) is required for efficient translation (McLaughlin *et al.*, 1981). Although the sequence at the 3' end of the 16S rRNA of these Gram-positive genera is identical to that of *E. coli* (McLaughlin *et al.*, 1981), genes from *E. coli* are poorly expressed in *B. subtilis*. It has been proposed that the principal obstruction to the expression of *E. coli* genes in *B. subtilis* occurs at the level of translation, as the SD sequences of *E. coli* have significantly lower complementarity to the 3' end of 16S rRNA than do those of *Bacillus* and *Staphylococcus* (Ehrlich and Sgarbetta, 1978).

The inability of *B. subtilis* to translate *E. coli* gene starts has been correlated with the absence of the largest 30S ribosomal protein found in *E. coli*, protein S1 (Higo *et al.*, 1982). Recently, the removal of the S1 protein from *E. coli* ribosomes was shown to result in the preferential translation of mRNA from *Bacillus* (Roberts and Rabinowitz, 1989). It was concluded that the S1 protein permits ribosomes to initiate translation at codons which are preceded by Shine-Dalgarno sequences with low complementarity to the 3' end of 16S rRNA.

Comparisons of the sequences preceding the translation start codons of genes from *Streptomyces* has indicated that the binding of ribosomes to *Streptomyces* mRNA does not require the degree of complementarity required apparently in *Bacillus* and *Staphylococcus* (Hopwood *et al.*, 1986; Seno and Baltz, 1988; this work). Unlike *Bacillus* and *Staphylococcus*, the 30S subunits of *Streptomyces* ribosomes contain a S1 protein (Roberts and Rabinowitz, 1988). In these respects, the initiation of translation in *Streptomyces* is analogous to the process in *E. coli*.

Although, the SD sequence can promote the initiation of translation, it is *not* essential in all cases. A number of *E. coli* genes lack SD sequences and yet are translated efficiently. Transcripts of the lambda *cI* gene which originate from the P<sub>M</sub> promoter, for example, lack a SD sequence and have the A of the AUG initiation codon at their 5' terminus. An examination of the sequences following the translation initiation codons of *E. coli* genes has shown that nearly all contain, at least, three to four conserved nucleotides that are complementary to some part of the 5'-AAAUUGAAGAGUUUGA- sequence located at the 5' terminus of 16S rRNA (Petersen *et al.*, 1988).

It has been proposed that sequences at the 5' end of 16S rRNA may be involved in mRNA recognition through a mechanism analogous to the well established "Shine and Dalgarno" interaction. In support, *cI* transcripts that lack SD sequences contain nucleotides downstream from the start codon which are complementary to a discrete region at the 5' end of the *E. coli* 16S rRNA (Petersen *et al.*, 1988).

At present, many laboratories interested in *Streptomyces* are sequencing genes involved either in morphological differentiation or in secondary metabolism. Using the FRAME program (Bibb *et al.*, 1984), it is usually relatively straightforward to identify the approximate boundaries of potential protein coding regions (PPCR). FRAME analyses is unable, however, to determine the precise bases at which a PPCR starts, particularly when two or more in-frame start codons are closely spaced (for further description, see Section 4.2.3.1). It has not always been possible to assign unambiguously translation initiation codons on the basis of complementarity to the 3' sequence of 16S rRNA.

For *E. coli* genes, the statistical analysis of sequences "downstream" from the proposed translation initiation codons (Petersen *et al.*, 1988) provided additional criteria to distinguish *bona fide* translation initiation codons from "false" starts in uncharacterised sequence data. Additional criteria would also aid the identification of translation initiation codons in *Streptomyces*. For simplicity, the term "gene start" (used by Petersen *et al.*, 1988) will be used to describe the region of mRNA, which is thought to be involved in the initiation of translation, at the start of a coding sequence.

The objective of this preliminary study was to establish if sequences downstream from the putative initiation codons of a selection of *Streptomyces* genes could form complementary base pairs with sequence at the 5' end of 16S rRNA and could be involved, in ribosome binding. The majority of the tentative gene starts which were examined were taken from the oxytetracycline cluster of *S. rimosus*.

## 7.2 Results and Discussion

### 7.2.1 Comparison of 16S rRNA Sequences from *Streptomyces* and *E. coli*

16S rRNAs from *S. coelicolor* and *S. lividans* have been sequenced (Baylis and Bibb, 1987; Suzuki and Yamata, 1987) and show greater than 99.5% identity. The 5' and 3' sequences of the *Streptomyces* 16S rRNA, corresponding to the *E. coli* sequences implicated in the binding of

the ribosome to mRNA, are shown in Figure 7.1, Panel A. The 3' sequences of the *Streptomyces* 16S rRNAs are almost identical to the equivalent *E. coli* sequence involved in Shine-Dalgarno base pairing. This is consistent with the finding that most of the *Streptomyces* genes which have been sequenced possess a reasonable Shine-Dalgarno sequence between 1 and 18 nucleotides before the start codon (see Table 7.1 for examples).

Alignment of the 16S rRNA sequences from *Streptomyces* revealed a slight difference in the 5' end points of *rrnD* from *S. coelicolor* and *rrnB* from *S. lividans*. The 5' termini of these 16S rRNAs may actually be different, alternatively, experimental error could be the cause of the slight discrepancy in their start points. For clarity, the 5' end of 16S rRNA from *S. lividans*, which is two nucleotides longer than the 5' end of 16S rRNA from *S. coelicolor* and one nucleotide longer than the *E. coli* 16S rRNA, was used as the reference sequence in the following comparisons. It was assumed that sequences at the extreme 5' end of 16S rRNA would be highly conserved in all *Streptomyces* spp.

The sequence of the 5' end of *Streptomyces* 16S rRNA only shows identity with nucleotides 8 to 16 of the *E. coli* sequence. The first 7 nucleotides of the *E. coli* 16S rRNA, which are implicated in "downstream" recognition, have no identity to the corresponding sequence in *Streptomyces*. The difference in sequence at the extreme 5' end of 16S rRNA may reflect the disparity in the preferred codon usage of *Streptomyces* and *E. coli*, as any interaction involving 16S rRNA within the coding sequence would be expected to be compatible with the coding requirements of the mRNA.

### 7.2.2 Complementarity Between Downstream Sequences and the 5' End of 16S rRNA

From the work of Petersen and co-workers (1988), which inspired this particular investigation, only the first 16nt of the 16S rRNA from *E. coli* were implicated in complementary base pairing with "downstream" sequences. Correspondingly, only the first 17nt of 16S rRNA from *Streptomyces* were analysed. In *E. coli*, nucleotides capable of forming



(A)

```

3'   AUUCCUCCA-(N)1516-AGUUUGAGAAGUAAAA 5' rrnB E.coli
      ++++++++          ++++++++ +
3'   UCUUCCUCCA-(N)1500-AGUUUGAGAGGCACU 5' rrnD S. coelicolor
      ++++++++          ++++++++*****
3'   UCUUCCUCCA-(N)1502-AGUUUGAGAGGCACUUA 5' rrnB S. lividans

```

(B)

```

          I---I      otrA
        I---I      otcC+
        I--I       otcX1
       I-----I    otcX2
       I-----I    otcZ

      I--I          actIII
     I-----I      c (øC31)
          I---I      dagA
          I-I        dhq+
           I--I      gra ORF1
          I- --I     gra ORF2
          I--I      pgm+
              I-I    tcmIa ORF1
          I--I      tcmIa ORF2
       I---I        whiG*1

```

Figure 7.1. (A) Alignment of the 5' and 3' sequence of 16S rRNA from *S. coelicolor* (Baylis and Bibb, 1987), *S. lividans* Suzuki and Yamata, 1987) and *E. coli* (Brosius *et al.*, 1981). A cross (+) indicates a nucleotide conserved between the three 16S rRNA sequences and the asterisks (\*) shows a nucleotide conserved between only the *Streptomyces* sequences. (B) Alignment of the 5' end of 16S rRNA from *Streptomyces* with complementary sequences in the tentative "downstream" regions of *otrA* (Doyle, 1987), *otcC* (Butler, unpublished work), *otcX* ORF1 and ORF2 (this work), *otcZ* (this work), *actIII* (Hallam *et al.*, 1988), øC31 *c* (Sinclair and Bibb, 1988), *dagA* (Buttner *et al.*, 1987), *dhq* (White, unpublished work), *gra* ORF1 and ORF2 (Sherman *et al.*, 1989), *pgm* (White, unpublished result), *tcmI* ORF1 and ORF2 (Bibb *et al.*, 1989) and *whiG* (Chater *et al.*, 1989). The line (I--I) indicates the extent of the proposed complementary base pairing, using the criteria described in Table 7.1.

**Table 7.1** Selection of gene starts from *Streptomyces* spp that show some degree of complementarity to a discrete region at the 5' end of 16S rRNA.

<b>A</b>	
<i>otrA</i>	ggaauccGAAcccccucc.aug.aacaagcUGAAUcuggggcauccuggcc
<i>otcC</i> <sup>+</sup>	uugaucagAGGAGcgugu.aug.cgguacgaCGUGgugaucgccggcgcg
<i>otcX1</i>	ggaccccGAGGAcguuc.aug.CCGUucacccagaaggaaucaccuac
<i>otcX2</i>	cgaggacGGAGGacggcg.aug.agggaacugcccgaauUCCGUGggau
<i>otcZ</i>	cccagaccAAGGAGuccc.aug.accaccacaccgcucgcccCGUGgccc
<b>B</b>	
<i>actIII</i>	aggcagGGAGGgGAacac.aug.gccacgcaggacUCCGaagucgcacug
<i>c</i> (øC31)	caacGAAGGGcgggacc.aug.aagcgggucACUCUCggcgggcggaag
<i>dagA</i>	cgaagAAGGAGaacgauc.gug.gucaaccgaCGUGAucucaucaagugg
<i>dhq</i> <sup>+</sup>	ucacggaccGGAGGaacc.gug.ccCCGcagccuggccaacgccccgauc
<i>gra</i> ORF1	caguacAGGAGGaccacg.gug.acccgacgcguaGUGAucacggggguc
<i>gra</i> ORF2	ccgcccggAGGAGGccgc.gug.agcacCCcUGAccgcccggcgggccguc
ORF1215* <sup>1</sup>	cucaccuGAAAGGcgcg.aug.uucgacaccgaagaCGUGggcguguuc
ORF1215* <sup>2</sup>	cacugacGAGGccggcac.gug.caacucaccUGAAaggggcgcgauuuc
<i>pgm</i> <sup>+</sup>	uggauuacgcucGgAAGc.aug.gccgacgcaCCGUacaagcugauccuc
<i>tcmI</i> ORF1	cgaaggugGGAGGcugag.aug.accCGGcacgcggagaagcgggguggug
<i>tcmI</i> ORF2	cccgaauuuccgugGGAu.gug.cugcugCCGUacaagaagcgcgcccgcg
<i>whiG</i> <sup>*1</sup>	aacggcacuAGGcGAcga.aug.ccccagcacacCUCCGggtccgaccgg
<i>whiG</i> <sup>*2</sup>	gcccgcgacGGUGgcagc.gug.cgaccgcCCGccccctcgacgcucgac
<b>C</b>	
<i>afsA</i> <sup>*2</sup>	auggacgcGGAGGccgag.gug.gugcatcCCGUGgggaucgagauggug
<i>afsA</i> <sup>*1</sup>	aug.gacgcggaggccgagguggugcatcccgugggggaucgag .....CCGorGUGg..or...CCGUGgg.....
<i>sta</i> <sup>+</sup>	aug.accacgacccauggcagcacguacgaauuccgcagcgcg .....UGg..or.CGU..or..uUCCG.....
<i>aph</i>	aug.gacgacagcacguugcgccggaaguacccgcaccacgag .....CGU.or.CCG..or...CCG.....

The Shine-Dalgarno (SD) and "downstream" sequences are shown in upper case. Nucleotides that extend the SD and downstream sequence if G:U base pairing is allowed are underlined. One mismatch was permitted in each sequence. The symbol (+) indicates translation start points confirmed by NH<sub>2</sub>-sequencing of the gene product. The symbol (\*1) indicates the authors first choice for the start codon. Likewise (\*2) indicates an alternative translation start point. In examples, where only three consecutive were identified, only the first triplet occurring downstream of the start codon is indicated. For references to sequences see text and Figure 7.1.

favourable interactions with 5' sequences from 16S rRNA were distributed between 4 to 21 nucleotides downstream of the start codon.

In this study, nucleotides as far as 39nt downstream of the proposed translation initiation codons of *Streptomyces* genes (+39) were analysed initially for complementarity. Similar to the situation in *E. coli*, it would appear that the coding regions of *Streptomyces* genes also contain a sequence of three or more contiguous nucleotides downstream of the translation initiation codon which are complementary to the 5' end of *Streptomyces* 16S rRNA (Table 7.1). Of the oxytetracycline genes, which have been sequenced completely (Table 7.1, Panel A), the putative translation initiation codons were designated initially after consideration of codon usage (Bibb *et al.*, 1984) and potential SD sites (Shine and Dalgarno, 1974). In support, all of these potential gene starts contain at least four consecutive nucleotides downstream of the start codons, which are complementary to the 5' end of 16S rRNA.

Sequences in *otcX* PPCR1 (UCCGUGgg) and *otcZ* (CCGUGg) have been identified, between +21 and +28 nucleotides, which could form highly favourable interactions with nucleotides at the 5' end of *Streptomyces* 16S rRNA. On the basis of this finding, it is suggested that putative "downstream" interactions should be considered as far downstream as +28 in *Streptomyces* genes. In this study, all of the gene starts which were examined contained at least three consecutive nucleotides downstream of the proposed start codon that were complementary to the 5' end of 16S rRNA.

Although, the most probable translation initiation codons for the oxytetracycline genes were assigned relatively easily, without consideration of "downstream" sequences, it has not always been possible to do so solely on the basis of suitably spaced Shine-Dalgarno sequences. The translation start codon for *whiG*, for example, could not be assigned unambiguously, as two potential start codons were preceded by suitably spaced sequences that had complementarity to the 3' end of 16S rRNA (*whiG*<sup>\*1</sup> and *whiG*<sup>\*2</sup>, Panel B) (Chater *et al.*, 1989). Analysis of these potential gene starts identified five consecutive nucleotides in

*whiG*<sup>\*1</sup> and three consecutive nucleotides in *whiG*<sup>\*2</sup>, respectively, which were complementary to the 5' end of 16S rRNA. Although, the *whiG*<sup>\*1</sup> gene start is predicted to form the most favourable interaction with the 30S ribosomal subunit, it remains a formal possibility that the *whiG*<sup>\*2</sup> start could also be a site for the initiation of translation.

The proposed gene start for ORF1215 of the element IS110 (Bruton and Chater, 1987) contains 4 consecutive nucleotides (6, if G:U base pairing is considered) downstream of the tentative AUG start with complementarity to the 5' end of 16S rRNA (Table 7.1. Panel B). Although, the tentative start codon is preceded by a region with complementarity to the 3' end of *Streptomyces* 16S rRNA, the precise region of complementarity is unconventional. It was noticed, however, that an in-frame GUG codon, 24 nucleotides upstream of the proposed AUG codon, is preceded by a reasonable SD-like sequences and contains four consecutive nucleotides downstream of this codon that are capable of forming complementary base pairing. On the basis of the overall complementarity to the 5' and 3' ends of 16S rRNA, it is suggested that the GUG codon upstream of the AUG codon should be considered as a potential start codon. This proposal is consistent with the FRAME plots (Bruton and Chater, 1987) which start to diverge at this point in the sequence.

As in *E. coli*, it is not uncommon in *Streptomyces* for the first nucleotide of the translation start codon to be also the 5' terminus of the mRNA, for example; the streptothricin acetyl transferase (*sta*) gene from *S. lavendulae* (Horinouchi *et al.*, 1987), the aminoglycoside phosphotransferase (*aph*) gene from *S. fradiae* (Janssen *et al.*, 1989) and the gene responsible for A-factor biosynthesis (*afsA*) of *S. griseus* (Horinouchi *et al.*, 1989). Analysis of the regions downstream of the proposed translation start codons for these genes has identified in each case, at least, three consecutive nucleotides which could form complementary base pairs with the 5' end of 16S rRNA (Table 7.1, Panel C). Interestingly, in the *afsA* message there is a GUG, located six codons downstream from the AUG codon at the 5' end. This GUG codon is preceded by five consecutive nucleotides that could form

conventional base pairs with the 3' end of 16S rRNA (*afsA*<sup>\*2</sup>) (Horinouchi *et al.*, 1989). Downstream of the GUG codon are 5 consecutive nucleotides (if G:U base pairing is considered) which could base pair with the 5' end of 16S rRNA. On the basis of complementarity of surrounding sequences to the 16S rRNA, the GUG codon might be expected to be the primary codon at which translation of the *afsA* message is initiated (*afsA*<sup>\*2</sup>). Site-directed mutagenesis has been used to create a frame shift between the AUG codon (that is the 5' codon of the *afsA* message) and the GUG codon (Horinouchi *et al.*, 1989). It was inferred from the ten-fold decrease in the expression of *afsA* in the mutant, that the AUG codon was the primary start codon, but, that translation at a low level could be initiated at the downstream GUG codon (Horinouchi *et al.*, 1989). An alternative hypothesis is that the GUG gene start, which has extensive complementarity to both the 5' and 3' end of 16S rRNA, was the primary site for translation initiation and mutagenesis upstream of the GUG codon severely affected the mRNA-ribosome interaction in an unknown manner. Amino-terminal sequencing of the *afsA* gene product would resolve these two possibilities.

It has been proposed that the lack of SD sequences in the mRNA of the above examples may be related to an unknown mechanism of differential gene expression (Horinouchi *et al.*, 1989). The "downstream" sequences of the above gene starts do not appear to contain significantly longer runs of consecutive nucleotides with complementarity to the 5' end of 16S rRNA, compared to those with SD sequences. In these examples, it is possible that other unknown factors are required to target 30S ribosomal subunits towards the translation codon at the extreme 5' end of the message. Although this preliminary analysis has focussed solely on the sequence content of gene starts, it should be noted that the secondary structure of mRNA around the start codon is also believed to have an important role in controlling translation. The reader is directed to papers by Iserentant and Fiers (1980) and de Smit and van Duin (1990) for discussion and references on the role of secondary structure in translation initiation.

### 7.2.3 Nucleotides of 16S rRNA implicated in binding to "downstream" regions

Alignment of the tentative "downstream" regions with the complementary sequences in the 5' end of 16S rRNA, revealed that only 7nt (AGGCACU) of the 17nt at the 5' end of *Streptomyces* 16S rRNA were repeatedly implicated in downstream-30S ribosomal subunit interactions (Table 7.1). Stated more precisely, the sequences of mRNA complementary to the 16S rRNA contain, in part, several conserved nucleotides from the string T.C<sub>or</sub>T.C<sub>or</sub>T.G.T.G.A<sub>or</sub>G within the first 25 nucleotides of the coding sequence immediately following the initiation codon. The alternative nucleotides within the string arise from possible G:U base pairing.

The nucleotides AGGCACU at the 5' end of 16S rRNA from *Streptomyces*, which are tentatively involved in mRNA recognition are not present at the equivalent position in 16S rRNA from *E. coli*. This is not surprising, perhaps, as the mol% G+C composition of these two bacteria are very different. It is conceivable that sequences at the 5' end of 16S rRNA, which are involved in "downstream" recognition, have evolved to accommodate the particular codon usage of individual bacterial genera. It would be interesting, therefore, to compare sequences at the 5' end of 16S rRNA and gene starts from a variety of genera with markedly different G+C compositions, to establish if there is any correlation with the biased usage of synonymous codons.

Although many *Streptomyces* promoters are not transcribed in *E. coli*, the binding of *Streptomyces* mRNA to *E. coli* ribosomes appears to proceed normally when transcription is facilitated by an adjacent *E. coli* promoter (Hunter and Baumberg, 1989). For example, the *otrAp1* promoter and the 5' end of the *otrA* resistance gene (nucleotides +1 to +1146) can direct expression of a *OtrA*:*LacZ* fusion protein in *E. coli* (McDowall, unpublished result). The expression of *Streptomyces* genes in *E. coli* (including examples where an *E. coli* promoter is used to direct transcription) suggests that "downstream" sequences in *Streptomyces* gene starts, which only rarely show any complementarity to the 5' end

of *E. coli* 16S rRNA, are not the principal determinants that direct the 30S ribosomal subunit to the correct translation initiation codon on mRNA. In general, an interaction with the Shine-Dalgarno sequence almost certainly directs the 30S ribosomal subunit to *bona fide* translation initiation codons (Gold and Stormo, 1987). Nevertheless, the statistical analysis of *E. coli* genes has revealed that contiguous nucleotides in the "downstream" region with complementarity to the 5' end of 16S rRNA are a conserved feature of authentic translation initiation codons. Downstream interactions may have an important role, therefore, in determining the efficiency of translation, perhaps by stabilising mRNA-30S subunit interactions during the formation of the initiation complex.

### 7.3 Scope of Future Studies

This limited study has shown that consecutive nucleotides downstream of the translation start codons in *Streptomyces* gene starts have the potential to bind the 5' end of 16S rRNA. This proposed interaction is envisaged to be important in determining the efficiency of translation.

In *E. coli*, the probabilities of contiguous nucleotides located downstream of translation initiation codons occurring at random with complementarity to sequences at the 5' end of 16S rRNA were calculated using a simple equation (Petersen *et al.*, 1988). It was assumed that the four nucleotides A, C, G and T were distributed randomly at the beginning of the coding regions. This would clearly not be a valid assumption for the analysis of "downstream" interactions in *Streptomyces*. The average mol G+C composition in the first, second and third position of *Streptomyces* codons is 70%, 45% and 90% respectively: this is clearly non-random.

The second approach, taken by Petersen and colleagues to analyse "downstream interactions" statistically, could be adopted to analyse *Streptomyces* gene starts. They created a file of "false" gene starts by searching the EMBL database for SD-like sequences followed

by suitably spaced AUG codons, which were not thought to be authentic sequences for the initiation of translation. The distribution of sequences complementary to the 5' end of 16S rRNA in files of approximately 150 "true" and "false" gene starts were then compared. In this way, the occurrence of complementary sequences was shown to be statistically significant in the downstream regions of real start codons. This analysis implied that the codon usage at the start of *E. coli* genes was biased to accommodate favourable base pairing with the 16S rRNA. The usage of synonymous codons in *Streptomyces* is already strongly biased towards those containing G and C residues. The nucleotides GGCACU at the 5' end of 16S rRNA, which are the principal nucleotides implicated in the base pairing with the downstream sequences (involved in greater than 50% of the proposed interactions, Figure 7.1) are representative of the average mol% G+C composition of the *Streptomyces* genome. It is possible, therefore, that the general biased usage of synonymous codons in *Streptomyces*, contributes to the high frequency with which consecutive nucleotides, complementary to the 5' end of 16S rRNA, are found downstream of translation start codons. With an increasing number of genes from *Streptomyces* being sequenced, currently around 50 have been published, it should soon be possible to analyse *Streptomyces* gene starts statistically, in a similar manner to *E. coli*, and determine if the observed complementarity in downstream regions of *Streptomyces* genes could be used as an additional criterion to elucidate authentic translation start points.



**CHAPTER 8**  
**CONCLUDING REMARKS**

## CONCLUDING REMARKS

The primary objective of the work described in this thesis was to characterise the transcription of the *otrA* resistance gene during the fermentation of oxytetracycline (OTC). The *otrA* gene was transcribed from two promoters, *otrAp1* and *otcCp1*, which were regulated differentially. During the period of vegetative growth that preceded antibiotic production, *otrA* was transcribed principally from the *otrAp1* promoter, which is located in the intergenic region between the 5' end of *otrA* and the 3' end of the biosynthetic gene *otcZ*. The relative amount of transcripts originating from *otrAp1* decreased within the total RNA population as the growth rate slowed. At the onset of antibiotic production, the *otrA* resistance gene was shown to be transcribed from beyond the 3' end of *otcZ*. A promoter was located upstream of *otcC*, which is the biosynthetic gene situated upstream of *otcZ*. The *otcCp1* promoter showed maximal activity during the early phase of OTC production. As S1 nuclease mapping experiments did not identify transcripts originating from immediately upstream of *otcZ*, it was concluded that during antibiotic production *otrA* was transcribed probably as part of a *otcC-otcZ-otrA* polycistronic message. Such an arrangement would ensure that the level of resistance to OTC increases in accordance with the level of antibiotic production. It remains to be shown directly, however, that a 5.3kb transcript extends from the promoter *otcCp1* to the rho-independent terminator, which is located immediately downstream of *otrA*.

Northern analysis (Alwine *et al.*, 1977; Fourney *et al.*, 1988) may appear to be a straightforward technique for confirming the presence of the 5.3kb *otcC-otcZ-otrA* polycistronic message. For unknown reasons, however, a reliable protocol has not been developed for analysing RNA from *Streptomyces* by Northern analysis. Recently, an alternative approach has been developed to map transcripts from *Streptomyces* at low resolution (Smith, unpublished work). Samples of total RNA are hybridized (as described for high-resolution S1 mapping) to a series of single-stranded M13 constructs, which contain overlapping inserts from the region of interest. The RNA samples are then digested with S1

nuclease. The RNA/DNA duplexes, which are not digested by S1 nuclease, are then resolved on agarose gels by electrophoresis. The RNA-protected fragments are then transferred to nylon membranes by capillary blotting under alkali conditions (as for Southern analysis). The immobilised DNA can then be analysed using radio-labelled oligonucleotides or probes produced by the random primer method. It is possible to map precisely the position of transcription units, by analysing the RNA-protected fragments with a number of different probes. The most time consuming part of this technique is the insertion of the DNA fragments from the region of interest into M13mp18 and M13mp19. Already much of the oxytetracycline cluster has been subcloned into these vectors to enable sequencing (Butler, unpublished work) and this technique may be useful for determining the organisation of transcription units within the *otc* cluster.

Sequencing of the *Sma*I<sub>9</sub>-*Bcl*I<sub>5</sub> fragment identified a single PPCR between the 3' end of *otcC* and the 5' end of *otrA*. The predicted amino acid sequence of the derived gene product was very similar to the amino acid sequence of a methyltransferase from bovine pineal glands (Ishida *et al.*, 1987) and a polypeptide predicted from a PPCR in a segment of the tetracenomycin cluster from *S. glaucescens*, which complements *tcmII* and *tcmV* mutants that accumulate C-3 demethyl intermediates (Hutchinson, pers. comm.). S-adenosylmethionine (SAM) has been shown to be the methyl group donor in the methylation reaction catalysed by bovine pineal gland. A possible SAM binding site was identified in a region which was highly conserved in all three polypeptides. It was concluded that the *otcZ* gene encoded a polypeptide that catalysed the C-6 methylation of the tetracyclic nucleus to form 6-methylpretetramid. It is predicted that *otcZ* mutants will accumulate C-6 demethyl intermediates. The *otcC* gene, which encodes the ATC oxygenase, was not described in any detail within this thesis, as it was sequenced in collaboration with M.J. Butler, formerly of Pfizer.

Two PPCRs, which were encoded on the opposite strand to *otcC*, *otcZ* and *otrA*, were identified within the *Sph*I<sub>12</sub>-*Sst*I<sub>13</sub> fragment of the *otcX* region. Comparison of the amino acid sequences of the predicted

products from these PPCRs with entries in the NBRF-protein and EMBL databases did not identify any similarity to polypeptides with known functions. The *SphI*-*SstI* fragment has been shown to hybridize to a region within the *actVa* region of the actinorhodin cluster of *S. coelicolor* (Butler *et al.*, 1990), which has been implicated in hydroxylation (Hopwood *et al.*, 1985; Cole *et al.*, 1987). The PPCRs in the *SphI*-*SstI* fragment may encode polypeptides with functions required for the C-4 hydroxylation of 6-MPT. Should this be the case, mutants altered in the functions of these polypeptides would be predicted to accumulate C-4 dehydroxyl intermediates.

Sequencing of DNA regions in the *otc* cluster, which hybridize with the *actI* and *actIII* probes (described by Malpartida *et al.*, 1987), has identified a number of PPCRs that are believed to encode polypeptides required for the biosynthesis of the polyketide backbone (Sherman, pers. comm.; Thamchaipenet, pers. comm.) Using the sequence data available for the *otc* cluster, it is now possible to introduce defined mutations into a number of PPCRs. Site-directed mutagenesis could be used to create a series of in-frame mutations, which would hopefully inactivate the gene products without disrupting transcription and translation. The mutated PPCR and flanking sequences would then be introduced into *S. rimosus* on derivatives of the temperature-sensitive vector pSG5 (Muth *et al.*, 1989), which contain resistance markers that can be selected in *S. rimosus* (Hunter, unpublished work). Replacement of the wild-type gene with the mutated gene would be achieved in two steps. It should be possible to isolate recombinants produced by a single recombination event between sequences cloned on the plasmid and homologous sequences on the chromosome. This would be achieved by incubating transformants at temperatures which restrict the replication of the plasmid, while still maintaining selection for plasmid markers. After culture of these recombinants at the restrictive temperature, isolates would be screened for the loss of the antibiotic resistance conferred by the plasmid and the reciprocal recombination event that removes the wild-type gene. A collaboration has been established with Dr R.A. Hill from the Chemistry Department at the University of Glasgow to analyse the polyketide products and shunt

metabolites produced by defined mutants. By combining expertise in analysing polyketide chemistry and the molecular genetics of *S. rimosus*, it should be possible to establish or confirm the function of individual genes in OTC biosynthesis.

Comparison of the promoters *otcCp1* and *otcXp1*, both of which are most active at the onset of antibiotic production, identified a conserved sequence that resembled *pho* boxes in *E. coli* (Makino *et al.*, 1986; 1988). This sequence has been designated the oxytetracycline *pho*-like box (*opl* box). Similar motifs have been identified in other promoters from *Streptomyces*, which are regulated by phosphate (Liras *et al.*, 1990). This conserved sequence is believed to be the binding site for a regulatory protein which activates transcription from promoters under conditions of low phosphate.

The isolation and characterisation of the regulatory protein, which is believed to bind *opl* boxes, represents the next major challenge in understanding the molecular mechanism that regulates OTC production. The isolation of total RNA from cultures of *S. rimosus* M15883 at different stages during the fermentation of OTC has shown that the *otcC-otcZ-otrA* polycistronic message and possibly other transcripts of the oxytetracycline cluster are most abundant at the onset of OTC production.

The regulatory protein which is proposed to bind promoters in the *otc* cluster and activate transcription is predicted to be most abundant and/or active early in the production of OTC. One of the advantages of studying antibiotic production in *S. rimosus* M4018 and M15883, which have been studied intensively due to the commercial importance of oxytetracycline, is that reproducible fermentations have been established which yield high levels of OTC. The ability to predict when OTC will be produced in batch cultures should ease the isolation of reasonable amounts of active regulatory protein, which in turn will aid the characterisation of *opl*-protein interactions.

Possibly, the most perplexing result of this work was the finding that the DNA elements which directed transcription of *xylE* in pIJ2843

constructs and hybridized to the *Sph*I<sub>12</sub>-*Sst*I<sub>13</sub> region did not correspond to either *otcCp1* or *otcXp1*. Should the plasmid-active-promoter (PAP) elements, which were cloned in *S. lividans* transformants 8(12) and 15, be active in the chromosome, they would direct the synthesis of RNA that was complementary to the mRNA for *otcX* PPCR1 and PPCR2. It was concluded that these antisense RNA transcripts were unlikely to be mRNAs, as PPCRs were not identified on the opposite strand to *otcX* PPCR1 and PPCR2.

Using total RNA isolated from *S. rimosus* at different stages in the fermentation of OTC, S1 nuclease mapping experiments failed to detect transcripts arising from the regions cloned in isolates 8 and 15. Primer extension analysis could be used to establish the transcription start points of the PAP elements, by annealing the reverse primer for pIJ2843 to total RNA isolated from cultures of isolates 8 and 15 (at the earliest possible time point when *xylE* was being expressed) and extending the primer using reverse transcriptase. One possible explanation for why primer extension analysis of the transcripts originating from *otcCp1* did not detect transcripts originating from the PAP elements, is that they terminated before the annealing site of the primer. Should the primer extension analysis of total RNA from *S. lividans* transformants 8 and 15 identify transcripts with a reasonable distance (>40 nucleotides) between the 5' end and sequences derived from pIJ2843, it would be possible to use oligonucleotides that were complementary to sequences close to the 5' end of the transcripts to identify short transcripts originating from the PAP elements in total RNA isolated from *S. rimosus* by primer extension. It is also possible that the antisense transcripts in the *otcX* region were not detected, as the PAP elements are only active transiently. Not knowing the role of these transcripts, it is impossible to predict when they might be synthesized. An alternative approach to detecting transcription from the PAP elements may be to insert a promoterless copy of the *xylE* reporter gene into the chromosome downstream of the PAP elements. The *xylE* reporter gene would be integrated as proposed previously for replacing wild-type genes in the *otc* cluster with genes mutated by site-directed mutagenesis.

Further analysis of oxytetracycline production in *S. rimosus* using molecular genetics promises to reveal many interesting facets about the regulation of antibiotic biosynthesis and its integration into the complex life cycle of *Streptomyces*. Genetic manipulation will also be an important tool for accessing the contribution of genes within the *otc* cluster to the biosynthesis of oxytetracycline.

## **BIBLIOGRAPHY**



- Aharonwitz, Y. and Demain, A.L. (1978). Carbon catabolite regulation of cephalosporin production in *Streptomyces clavuligerus*. *Antimicrob. Agents Chemother.* 14, 159-164.
- Alwine, J.C., Kemp, D.J. and Stark, G.R. (1977). Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proc. Nat. Acad. Sci., USA*, 74, 5150-5355.
- Alves, T.P.T.S. (1989). "Development of plasmid vectors for regulated expression of cloned genes in *Streptomyces lividans* and *Escherichia coli*. PhD Thesis, University of Glasgow, Scotland.
- Anderson, M.L.M. and Young, B.D. (1985). Quantitative filter hybridization. In: "Nucleic Acid Hybridization: a practical approach." Hames, B.D. and Higgins, S.J. (eds). IRL Press. pp.73-111.
- Antonov, P.P., Ivanov, I.G., Markov, G.G. and Benigni, R. (1978). Reassociation analysis of DNA in studying the genome size of *Streptomyces*. *Studia Biophysica*, Berlin, 69, 67-74.
- Anzai, H., Murakami, T., Imai, S., Satoh, A., Nagaoka, K. & Thompson, C.J. (1987). Transcriptional regulation of bialaphos biosynthesis in *Streptomyces hygroscopicus*. *J. Bacteriol.* 169, 3482-3488.
- Arison, B.H. and Omura, S. (1974). Revised structure of cerulenin. *J. Antibiot.* 27, 28-30.
- Asturias, J.A., Liras, P. and Martin J.F. (1990). Phosphate control of *pabS* gene transcription during candicidin biosynthesis. *Gene* 93, 79-84.
- Axelrod, J. and Weissbach, H. (1961). Purification and properties of hydroxyindole O-methyl transferase. *J. Biol. Chem.* 236, 211-213.
- Baltz, R.H., Seno, E.T., Stonesifer, J., Matsushima, P. and Wild, G.M. (1981). Genetics and biochemistry of tylosin production by *S. fradiae*. In: "Microbiology 1981". Am. Soc. Microbiol. pp 371-375.
- Bartel, P.L., Chun-Bao, Z., Lampel, J.S., Dosch, D.C., Connors, N.C., Struhl, W.R., Beale Jr., J.M. and Floss, H.G. (1990). Biosynthesis of anthraquinones by interspecies cloning of actinorhodin biosynthetic genes in streptomycetes: clarification of actinorhodin gene functions. *J. Bacteriol.* 172, 4816-4826.
- Bascaran, V., Hardisson, C. and Brana, A.F. (1989a). Regulation of nitrogen catabolic enzymes in *Streptomyces clavuligerus*. *J. Gen. Microbiol.* 135, 2465-2474.
- Bascaran, V., Hardisson, C. and Brana, A.F. (1989b). Isolation and characterisation of nitrogen-deregulated mutants of *Streptomyces clavuligerus*. *J. Gen. Microbiol.* 135, 2475-2474.

- Bayer, E., Gugel, K.H., Hagele, K., Hagennaier, H., Jessipow, S., Konig, W.A. and Zahner, H. (1972). Stoffwechselprodukte von Mikroorganismen. Phosphinothricin und phosphinothricyl-alanyl-alanine. *Helv. Chem. Acta.* 55, 224-239.
- Baylis, H.A. and Bibb, M.J. (1987). The nucleotide sequence of 16S rRNA from *Streptomyces coelicolor* A3(2). *Nuc. Acid Res.* 15, 7176.
- Baylis, H.A. and Bibb, M.J. (1988). Transcriptional analysis of the 16S rRNA gene of the *rrnD* gene set of *S. coelicolor* A3(2). *Mol. Microbiol.* 2, 569-580.
- Behal, V. (1986). Enzymes of secondary metabolism: regulation of their expression and activity. In: "Regulation of Secondary Metabolite Formation". Kleinkawf, H., Duhren, H., Domauer, H., and Nesemann, G. (ed). VCH Verlagsgesellschaft, Weinheim pp. 264-281.
- Behal, V. (1987). The tetracycline fermentation and its regulation. *CRC Central Reviews in Biotechnology* 5, 275-318.
- Behal, V., Hostalek, Z. and Vanek, Z. (1979). Anhydrotetracycline oxygenase activity and biosynthesis of tetracyclines in *Streptomyces aureofaciens*. *Biotechnol. Lett.* 1, 177-193.
- Berk, A.J. and Sharp, P.A. (1977). Sizing and mapping of early adeno virus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. *Cell*, 12, 721-732.
- Bhatnager, R.K., Doull, J.L. and Vining, L.C. (1988). Role of the carbon source in regulating chloramphenicol production by *Streptomyces venezuelae*: studies in batch and continuous cultures. *Can. J. Microbiol.* 35, 1217-1223.
- Bibb, M.J. (1985). Nucleotide sequences encoding and promoting expression of three antibiotic resistance genes indigenous to *Streptomyces*. *Mol. Gen. Genet.* 199, 26-36.
- Bibb, M.J. and Cohen, S.N. (1982). Gene expression in *Streptomyces*: construction and application of promoter-probe plasmid vectors in *Streptomyces lividans*. *Mol. Gen. Genet.* 187, 265-277.
- Bibb, M.J. & Janssen, G.R. (1987). Unusual features of transcription and translation of antibiotic resistance genes in antibiotic producing *Streptomyces*. In: "Genetics of Industrial Microorganisms." Proceedings of the Fifth International Symposium. Alcevic, M., Hranueli, D. and Toman, A. (eds). Pliva.
- Bibb, M.J., Biro, S., Motamedi, H., Collins, J.F. and Hutchinson, C.R. (1989). Analysis of the nucleotide sequence of the *S. glaucescens* *tcml* genes provides key information about the enzymology of polyketide antibiotic biosynthesis. *EMBO J.* 8, 2727-2736.

- Binnie, C., Warren, M. & Butler, M.J. (1989). Cloning and heterologous expression in *Streptomyces lividans* of *Streptomyces rimosus* genes involved in oxytetracycline biosynthesis. *J. Bacteriol.* 171, 887-895.
- Birnboim, H.C. and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nuc. Acid Res.* 7, 1513-1523.
- Blum, R.H. and Carter, S.K. (1974). Adriamycin - a new anti-cancer drug with significant clinical activity. *Ann. Intern. Med.* 80, 249-259.
- Brawner, M., Taylor, D. and Fornwaid, J. (1990). Expression of the soluble CD-4 receptor in *Streptomyces*. In: "Molecular Biology of Streptomycetes". Bibb, M.J., Losick, R. and Taylor D. (ed). Abstracts of the UCLA Symposium on Molecular and Cellular Biology, Colorado, USA.
- Brosius, J., Dull, T.J., Sleeter, D.D. and Noller, H.F. (1981). Gene organisation and primary structure of a ribosomal RNA operon from *Escherichia coli*. *J. Mol. Biol.* 148, 107-127.
- Bruton, C.J. and Chater, K.F. (1987). Nucleotide sequence of IS110, an insertion sequence of *Streptomyces coelicolor* A3(2). *Nuc. Acid Res.* 15, 7053-7065.
- Burg, R.W., Brinton, M.M., Baker, E.E., Birnbaum, J., Currie, S.A., Hartman, R., Kong, Y-l., Monaghan, R.L., Olson, G., Putter, I., Tunac, J.B., Wallick, H., Striple, E.O., Oiwa, R. and Omura, S. (1979). Avermectins, new family of potent antihelmintic agents: producing organisms and fermentation. *Antimicro. Agents Chemo.* 15, 361-367.
- Butler, M.J., Friend, E.J., Hunter, I.S., Kaczmarek, F.S., Sugden, D.A. & Warren, M. (1989). Molecular cloning of resistance genes and architecture of a linked gene cluster involved in biosynthesis of oxytetracycline by *Streptomyces rimosus*. *Mol. Gen. Genet.* 215, 231-238.
- Butler, M.J., Binnie, C., Hunter, I.S., Sugden, D.A. and Warren, M. (1990). Genetic manipulation of the oxytetracycline biosynthetic pathway genes. In: "Developments in Industrial Microbiology." 31. Pierce, G.E. (ed). Society for Industrial Microbiology, pp. 41-50.
- Buttner, M.J., Chater, K.F. and Bibb, M.J. (1990). Cloning, disruption and transcriptional analysis of three RNA polymerase sigma factor genes of *Streptomyces coelicolor* A3(2). *J. Bacteriol.* 172, 3367-3378.
- Buttner, M.J., Fearnley, I.M. and Bibb, M.J. (1987). The agarase gene (*dagA*) of *Streptomyces coelicolor* A3(2): nucleotide sequence and transcriptional analysis. *Mol. Gen. Genet.* 209, 101-109.
- Buttner, M.J., Smith, A.M. & Bibb, M.J. (1988). At least three different RNA polymerase holoenzymes direct transcription of the agarase gene (*dagA*) of *Streptomyces coelicolor* A3(2). *Cell* 52, 599-607.

- Calzone, F.J., Britten, R.J. and Davidson, E.H. (1987). Mapping of gene transcripts by nuclease protection assays and cDNA primer extension. In: "Guide to Molecular Cloning Techniques." Berger, S.L. and Kimmel, A.R. (eds). Academic Press, London. pp. 611-632.
- Cami, B. and Kourilski, P. (1978). Screening of cloned recombinant DNA in bacteria by *in situ* colony hybridization. *Nuc. Acid Res.* 5, 2381-2390.
- Campbell, D. A. (1986). Similarities and differences in regulation of primary and secondary metabolism. In: "Abstracts of the 106<sup>th</sup> meeting of the Society for General Microbiology." University of Warwick.
- Cane, D.E., Celmer, W.D. and Westley, J.W. (1983). Unified stereochemical model of polyether antibiotic structure and biogenesis. *J. Am. Chem. Soc.* 105, 3594-3600.
- Casey, J. and Davidson, N. (1977). Rates of formation and thermal stabilities of RNA:DNA and DNA:DNA duplexes at high concentrations of formamide. *Nuc. Acid Res.* 4, 1539-1552.
- Cashel, M. (1969). The control of ribonucleic acid synthesis in *Escherichia coli*. IV. Relevance of unusual phosphorylated compounds from amino acid clone stringent strains. *J. Biol. Chem.* 244, 3133-3141.
- Cashel, M. (1975). Regulation of bacterial ppGpp and pppGpp. *Ann. Rev. Microbiol.* 29, 301-318.
- Cashel, M. and Rudd, K.E. (1987). Stringent Response. In *Escherichia coli* and *Salmonella typhimurium*, Cellular and Molecular Biology. Neidhardt, F.C., Ingraham, J.L., Brooks Low, K., Magasanik, B., Schaechter, M. and Umberger, H.E (eds). American Society for Microbiology, Washington, D.C. pp. 1410-1438
- Chater, K.F. (1984). Morphological and Physiological differentiation in *Streptomyces*. In: "Microbial Development." Losick, R. and Shapiro, L. (eds). Cold Spring Harbor Laboratory Press, New York. pp. 89-115.
- Chater, K.F. (1990). Multilevel regulation of *Streptomyces* differentiation. *TIG* 5, 372-377.
- Chater, K.F. & Bruton, C.J. (1983). Mutational cloning in *Streptomyces* and the isolation of antibiotic production genes. *Gene* 26, 67-78.
- Chater, K.F. and Bruton, C.J. (1985). Resistance, regulatory and production genes for the antibiotic methylenomycin are clustered. *EMBO Journal* 4, 1893-1897.
- Chater K.F. and Merrick, M.J. (1979). *Streptomyces*. In "Studies in Microbiology. Volume 1. Developmental Biology of Prokaryotes." Pavish, J.H. (ed). University of California Press (Berkley and Los Angeles). pp. 93-114.

- Chater, K.F., Bruton, C.J., Plaskitt, K.A., Buttner, M.J., Mendez, C. and Helmann, J.D. (1989). The developmental fate of *S. coelicolor* hyphae depends upon a gene product homologous with the motility sigma factor of *B. subtilis*. *Cell* 59, 133-143.
- Chen, C.W., Lin, H.F., Cuo, C.L., Tsai, H.L. & Tsai, J.F.Y. (1986). Cloning of a DNA sequence conferring cephamycin C production in a heterologous host. Abstract W7-P14. Fifth International Symposium on the Genetics of Industrial Microorganisms, Split, Yugoslavia, 1986. p.164.
- Chopra, I. (1988). Molecular mechanisms involved in the transport of antibiotics into bacteria. *Parasitology* 96, 525-544.
- Clayton, T.M. and Bibb, M.J. (1990). *Streptomyces* promoter probe plasmid that utilises the *xylE* gene from *Pseudomonas putida*. *Nuc. Acid Res.* 18, 1077.
- Close, T.J. and Rodriguez, R.L. (1982). Construction and characterisation of the chloramphenicol resistance gene cartridge: a new approach to the transcriptional mapping of extrachromosomal elements. *Gene* 20, 305-316.
- Cortes, J., Liras, P., Castro, J.M. and Martin, J.F. (1986). Glucose regulation of cephamycin biosynthesis in *Streptomyces lactamdurans* is exerted on the formation of  $\alpha$ -aminoadipyl-crysteinyl-valine and deacetoxycephalosporin C synthase. *J. Gen. Microbiol.* 132, 1805-1814.
- Cortes, J., Haydock, S.F., Roberts, G.A., Bevitt, D.J. and Leadlay, P.F. (1990). An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of *Saccharospora erythaea*. *Nature* 348, 176-178.
- Cowing, D.W., Bardwell, J.C.A., Craig, E., Woolford, C., Hendrix, R.W and Gross, C.A. (1985). Consensus sequence for *Escherichia coli* heat shock promoters. *Proc. Nat. Acad. Sci.* 82, 2679-2683.
- Craig, N.L. and Nash, H.A. (1984). *E. coli* integration host factor binds to specific sites in DNA. *Cell* 39, 707-710.
- Davis, N.K. and Chater, K.F. (1990). Spore colour in *Streptomyces coelicolor* A3(2) involves the developmentally regulated synthesis of a compound biosynthetically related to polyketide antibiotics. *Molec. Microbiol.* 4, 1679-1693.
- de Crombrughe, B., Busby, S. and Buc, H. (1984). Cyclic AMP receptor protein: role in transcription activation. *Science* 224, 831-838.
- de Smit, M.H. and van Duin, J. (1990). Secondary structure of the ribosome binding site determines translational efficiency. A quantitative approach. *Proc. Nat. Acad. Sci. USA* 87, 7668-7672.

- Delic, V., Pigac, J. and Sermonti, G. (1969). Detection and study of cosynthesis of tetracycline antibiotics by an agar method. *J. Gen. Microbiol.* 55, 103-108.
- Demsey, W.B. (1987). Transcript analysis of the plasmid R100 *traJ* and *finP* genes. *Mol. Gen. Genet.* 209, 533-544.
- Denhardt, D.T. (1966). A membrane-filter technique for the detection of complementary DNA. *Biochem. Biophys. Res. Com.* 23, 641-646.
- Devereux, J., Haeberli, P. and Smithies O. (1984). A comprehensive set of sequence analysis pprograms for the VAX. *Nuc. Acid res.* 12, 387-395.
- Distler, J., Ebert, A., Mansouri, K., Possowatzki, K., Stock, M. and Piepersberg, W. (1987). Gene cluster for streptomycin biosynthesis in *Streptomyces griseus*: nucleotide sequence of three genes and analysis of transcriptional activity. *Nuc. Acid Res.* 15, 8041-8056
- Distler, J., Mansouri, K., Possiwotzki, K., Heinzl, P., Mayer, G. and Pipersberg, W. (1988). Organisation and expression of streptomycin production genes in *S. griseus*. In: "Abstracts of the 108th Meeting of the Genetical Society", Norwich, UK.
- Doyle, D., Butler, M.J. & Hunter, I.S. (1988). Molecular analysis of the oxytetracycline resistance gene, *otrA*, from *Streptomyces rimosus*. *Heredity* 61, 305.
- Doyle, D., McDowall, K.J., Butler, M.J. and Hunter, I.S. (1991). Characterisation of an oxytetracycline gene, *otrA*, of *Streptomyces rimosus* nucleotide sequence, deduced function and promoter analysis. *Mol. Microbiol.* (submitted).
- Drlica, K. (1987). The nucleoid. In: "*Escherichia coli* and *Salmonella typhimurium*". Neidhardt, F.C., Ingraham, J.L., Brooks Low, K., Magasanik B., Schaechter, M. and Umberger, H.E (eds). American Society for Microbiology. pp. 91-103.
- Dunn, T.M., Hahn, S., Ogden, S. and Schleif, R.F. (1984). An operator at -280 base pairs that is required for repressor of *araBAD* operon promoter: addition of DNA helical turns between the operator and promoter cyclically hinders repression. *Proc. Nat. Acad. Sci. USA* 81, 5017-5020.
- Ehrlich, S.D. and Sgaramella, V. (1978). Barriers to heterospecific gene expression among prokaryotes. *Trends in Biochem. Sci.* 3, 259-261.
- Enquist, L.W. and Bradley, S.G. (1971). Characterisation of deoxyribonucleic acid from *Streptomyces venezuelae* species. *Dev. Ind. Microbiol.* 12, 2431-2441.
- Escallante, L., Lopez, H., Mateus, R., Lara, F. and Sanchez, S. (1982). Transient repression of erythromycin formation in *Streptomyces erythraeus*. *J. Gen. Microbiol.* 128, 2011-2015.

- Feinberg, A.P. and Vogelstein, B. (1983). A technique for radio-labelling DNA restriction exonuclease fragments to high specific activity. *Anal. Biochem.* 132, 6-13.
- Feinberg, A.P. and Vogelstein, B. (1984). Addendum to Feinberg and Vogelstein (1983). *Anal. Biochem.* 137, 266-267.
- Ferrari, E., Henner, A.J., Peregom, A.F. and Hoch, J.A. (1988). Regulation of post exponential gene expression: transcriptional control of the subtilisin gene. In: "Genetics and Biotechnology of Bacilli", Ganeson, A.T. and Hoch, J.A. (ed). Academic press Inc. pp 39-44.
- Fickett, J.W. (1982). Recognition of protein coding regions in DNA sequences. *Nuc. Acid Res.* 10, 5303-5318.
- Flashner, Y. and Gralla, J.D. (1988). Dual mechanism of repression of a distance in the *lac* operon. *Proc. natl. Acad. Sci. USA* 85, 8968-8972.
- Fourney, R.M., Miyakoshi, J., Day, R.S. and Paterson, M.C. (1988). Northern blotting: efficient RNA staining and transfer. *Bethesda Res. Lab. Focus* 10. 5-7.
- Galloway, J.L. and Platt, T. (1986). Control of prokaryotic gene expression by transcription termination. In: "Regulation of Gene Expression - 25 years on." Booth, I.R. and Higgs, C.F. (eds). Cambridge University Press. pp 155-178.
- Ganoza, M.C., Sullivan, P., Cunningham, C., Harder, P., Kofoed, E.C. and Neilson, T. (1982). Effect of bases contiguous to AUG on translation initiation. *J. Biol. Chem.* 257, 8228-8232.
- Gibson, T.J. (1984). "Studies on the Epstein-Barr virus genome". PhD Thesis, Cambridge University, England.
- Gil, J.A. and Hopwood, D.A. (1983). Cloning and expression of a p-aminobenzoic acid synthetase gene of the candicin producer *Streptomyces griseus*. *Gene* 25, 119-132.
- Gil, J.A., Naharó, G., Villanueva, J.R. & Martin, J.F. (1985). Characterization and regulation of p-aminobenzoic acid synthase from *Streptomyces griseus*. *J. Gen. Microbiol.* 131, 1279-1287.
- Gilbert, W., Majers, J. and Maxam, A. (1976). In: "Dahlem Workshop of Chromosomes". Abakon Verlagsgesellschaft, Berlin, pp. 167-178.
- Gold, L. and Stormo, G. (1987). Translation initiation. In: "*Escherichia coli* and *Salmonella typhimurium*." Neidhardt, F.C., Ingraham, J.L., Brooks, Low, K., Magasanik, B., Schaechter, M. and Umberger, H.E. (eds). Cellular and Molecular Biology, American Soc. for Microbiol., Washington DC. pp. 1302-1307.
- Gold, L., Pribnow, D., Schneider, T., Shinfeld, S., Swebiliun, B.S. and Stormo, G. (1981). Translational initiation in prokaryotes. *Ann. Rev. Microbiol.* 35, 365-403.

- Goodfellow, M., and O'Donnell, A. G. (1989). Search and discovery of industrially-significant actinomycetes. In: "Microbial products: New approaches." Symposium of the Society for General Microbiology. Baumberg, S., Hunter, I. S., and Rhodes, P. M. (eds). Cambridge University Press. pp.121-162.
- Grafe, U., Schade, W., Erritt, I. and Fleck, W.F. (1982). A new inducer of anthracycline biosynthesis from *Streptomyces viridochromogenes*. J. Antibiot. 35, 1722-1723.
- Grafe, U., Reinhardt, G., Schade, W., Erritt, I., Fleck, W.F. and Radks, L. (1983). Interspecific reducers of cytodifferentiation and anthracycline biosynthesis from *Streptomyces bikiniensis* and *S. cyaneofuscatus*. Biotechnol. Lett. 5, 591-596.
- Gralla, J.D. (1989). Bacterial gene regulation from distant DNA sites. Cell 57, 193-195.
- Gralla, J.D. (1990). Regulation of Transcription in Procaryotes. 117<sup>th</sup> Meeting of the Society for General Microbiology. University of Aberdeen, September 11-12, 1990.
- Granozzi, C., Biletia, R., Passantino, R., Sollazzo, M. and Puglia, A.M. (1990). A breakdown in macromolecular synthesis preceding differentiation in *Streptomyces coelicolor* A3(2). J. Gen. Microbiol. 136, 713-716.
- Grossman, A.D. and Losick, R. (1986). RNA polymerase heterogeneity in bacteria. In: "Regulation of Gene Expression - 25 years on." Booth, I.R. and Higgins, C.F. (eds). Cambridge Academic Press.
- Grossman, A.D., Erickson, J.W. and Gross, C.A. (1984). The *htpR* gene product of *E. coli* is a sigma factor for heat-shock promoters. Cell 38, 383-390.
- Guijaró, J., Santamaria, R., Schauer, A., and Losick, R. (1988). Promoter determining the timing and spatial localization of transcription of a cloned *Streptomyces coelicolor* gene encoding a spore-associated polypeptide. J. Bacteriol. 170, 1895-1901.
- Guthrie, E.P. (1989). Partial suppression of a pleiotropic antibiotic non-producing mutant of *Streptomyces coelicolor*. In: "Abstracts of the 113<sup>th</sup> meeting of the Society for General Microbiology." University of Cambridge.
- Guthrie, E.P. and Chater, K.F. (1990). The level of transcription required for production of a *Streptomyces coelicolor* antibiotic is conditionally dependent on a tRNA gene. J. Bacteriol. 172, 6139-6193.



- Guzman, P., Westpheling, J. and Youngman, P. (1988). Characterisation of the promoter region of the *Bacillus subtilis* *spoIIIE* operon. *J. Bacteriol.* 170, 1598-1609.
- Haber, R. and Adhya, S. (1988). Interaction of spatially separated protein-DNA complexes for control of gene expression: Operator conversions. *Proc. Natl. Acad. Sci. USA* 85, 9683-9687.
- Hall, M.N., Gabay, J., Debarbouille, M. and Schwartz, M. (1982). A role for mRNA secondary structure in the control of translation initiation. *Nature* 295, 616-618.
- Hallam, S.E., Malpartida, F. and Hopwood, D.A. (1988). Nucleotide sequence, transcription and deduced function of a gene involved in polyketide antibiotic synthesis in *Streptomyces coelicolor*. *Gene* 74, 305-320.
- Hamilton, E.P. and Lee, N. (1988). Three binding sites for AraC protein are required for autoregulation of *araC* in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 85, 1749-1753.
- Hanahan, D. (1983). Studies on transformation of *E. coli* with plasmids. *J. Mol. Biol.* 166, 557-580.
- Hara, O. and Beppu, T. (1982). Mutants blocked in streptomycin production in *Streptomyces griseus* - the role of A-factor. *J. Antibiot.* 35, 349-358.
- Hara, O. and Beppu, T. (1982). Induction of streptomycin inactivating enzyme by A-factor in *Streptomyces griseus*. *J. Antibiot.* 35, 1208-1215.
- Hara, O., Horinouchi, S., Uozumi, T. and Beppu, T. (1983). Genetic analysis of A-factor synthesis in *Streptomyces coelicolor* A3(2) and *Streptomyces griseus*. *J. Gen. Microbiol.* 129, 2939-2944.
- Hardisson, C., Manzanal, B., Mendez, C. and Brana, A. (1986). In: "Biological Biochemical and Biomedical Aspects of Actinomycetes." Szabo, G., Biro, S. and Goodfellow, M. (eds). Akademiai Kiado, Budapest. pp. 433-442.
- Harms, E., Higgins, E., Chen, J-W. and Umbarger, H.E. (1988). Translational coupling between the *ivlD* and *ivlA* genes of *Escherichia coli*. *J. Bacteriol.* 170, 4798-4807.
- Hawley, D.K. and McClure, W.R. (1983). Compilation and analysis of *E. coli* promoter DNA sequences. *Nuc. Acid Res.* 11, 2237-2255.
- Helmann, J.D. and Chamberlin, M.J. (1987). DNA sequence analysis suggests that the expression of flagellar and chemotaxis genes in *Escherichia coli* and *Salmonella typhimurium* is controlled by an alternative  $\sigma$  factor. *Proc. Natl. Acad. Sci., USA* 84, 6422-6424.

- Helmann, J.D., Marquez, L.M. and Chamberlin, M.J. (1988). Cloning, sequencing and disruption of the *Bacillus subtilis*  $\sigma^{28}$  gene. *J. Bacteriol.* 170, 1568-1574.
- Henner, D.J., Ferrari, E., Perego, M. and Hoch, J.A. (1988). Upstream activating sequences in *Bacillus subtilis*. In: "Genetics and Biotechnology of Bacilli" Ganesan, A.T. and Hoch, J.A. (ed). Academic Press pp. 3-10.
- Hershberger, C. (1989). Transcription of tylosin biosynthetic genes. In: "Microbes Under Stress: Metabolic and Developmental Choices." Abstract of the 31st Harden Conference of the Biochemical Society. Broda, P.M.A. and Cullum, J. (eds).
- Hess, J.F., Oosawa, K., Kaplan, N. and Simon M.I. (1988). Phosphorylation of three proteins in the signaling pathway of bacterial chemotaxis. *Cell*, 53, 79-87.
- Higgins, C.F., Dorman, C.J., Stirling, D.A., Waddell, L., Booth, I.R., Gerhard, M. and Bremer, E. (1988). A physiological role for DNA supercoiling in the osmotic regulation of gene expression in *S. typhimurium* and *E. coli*. *Cell* 52, 569-584.
- Higo, K., Otaka, E. and Osawa, S. (1982). Purification and characterization of 30S ribosomal proteins from *Bacillus subtilis* correlation to *Escherichia coli* 30S proteins. *Mol. Gen. Genet.* 185, 239-244.
- Hillen, W., Schollmeier, K. and Gatz, C. (1984). Control of expression of the Tn10-encoded tetracycline resistance operon. II. Interaction of RNA polymerase and TET repressor with the *tet* operon regulatory region. *J. Mol. Biol.* 172, 185-201.
- Hirschman, J., Wong, P.K., Sei, K., Keener, J. and Kustu, S. (1985). Products of nitrogen regulatory genes *ntrA* and *ntrC* of enteric bacteria activate *glnA* transcription *in vitro*: evidence that the *ntrA* product is a sigma factor. *Proc. Nat. Acad. Sci.* 82, 7525-7529.
- Hopwood, D.A. (1988). Towards an understanding of gene switching in *Streptomyces*, the basis of sporulation and antibiotic production. *Proc. R. Soc. London Ser. B* 235, 121-138.
- Hopwood, D.A. and Kieser, T. (1990). The *Streptomyces* genome. In: "The Bacterial Chromosome." Drlica, K. and Riley, M. (ed). American Society for Microbiology, Washington DC. pp. 147-162.
- Hopwood, D.A. and Sherman, D.H. (1990). Molecular genetics of polyketides and its comparison to fatty acid biosynthesis. *Ann. Rev. Genet.* 24, 37-66.
- Hopwood, D.A., Bibb, M.J., Chater, K.F., Kieser, T., Bruton, C.J., Kieser, H.M., Lydiate, D.J., Smith, C.P., Ward, J.M. and Schrempf, H. (1985a). In: "Genetic manipulation of *Streptomyces*: a laboratory manual." The John Innes Foundation.

- Hopwood, D.A., Malpartida, F. and Chater, K.F. (1985b). Gene cloning to analyse the organisation and expression of antibiotic biosynthesis genes in *Streptomyces*. In: "Regulation of Secondary Metabolite Formation". Kleinkauf, H., van Dohren, H., Dornan, H. and Nesemann, G. (eds). Weinheim, VCH Verlagsgesellschaft.
- Hopwood, D.A., Malpartida, F., Kieser, H.M., Ikeda, H., Duncan, J., Fujii, I., Rudd, B.A.M., Floss, H.G. and Omura, S. (1985b). Production of hybrid antibiotics by genetic engineering. *Nature* 314, 642-646.
- Hopwood, D.A., Bibb, M.J., Chater, K.F., Janssen, G.R., Malpartida, F. and Smith, C.P. (1986). Regulation of gene expression in antibiotic-producing *Streptomyces*. In: "Regulation of Gene Expression - 25 Years On." Booth, I.R. and Higgins, C.F. (eds). Cambridge University Press, pp 251-276.
- Horinouchi, S. and Beppu, T. (1985). Construction and application of a promoter-probe plasmid that allows chromogenic identification in *Streptomyces lividans*. *J. Bacteriol.* 162 (1), 406-412.
- Horinouchi, S. and Beppu, T. (1987). A-factor and regulatory network that links secondary metabolism with cell differentiation in *Streptomyces*. In: "Proceedings of the Fifth International Symposium on the Genetics of Industrial Microorganisms." Alacevic, M., Hranueli, D. and Toman Z. (eds). Pliva Zagreb.
- Horinouchi, S., Suzuki, H. and Beppu, T.C. (1986). Nucleotide sequence of *afsB*, a pleiotropic gene involved in secondary metabolism in *Streptomyces coelicolor* A3(2) and *Streptomyces lividans*. *J. Bacteriol.* 168, 207-269.
- Horinouchi, S., Furuya, K., Nishiyama, M., Suzuki, H. and Beppu, T. (1987). Nucleotide sequence of the Streptothricin acetyl transferase gene from *Streptomyces lavendulae* and its expression in heterologous hosts. *J. Bacteriol.* 169, 1929-1937.
- Horinouchi, S., Suzuki, H., Nishiyama, M. and Beppu, T. (1989). Nucleotide sequence and transcriptional analysis of the *Streptomyces griseus* gene (*afsA*) responsible for A-factor biosynthesis. *J. Bacteriol.* 171 (2), 1206-1210.
- Humphreys, G.O., Bailey, C.R., and McKillop, C.A. (1987). Isolation of genes for biosynthesis of polyketide antibiotics by screening a clone library for hybridisation with nucleic acid sequence coding for a second antibiotic. GB Patent Application 8616851.
- Hunt, T.P. and Magasanik, B. (1985). Transcription of *glnA* by purified *Escherichia coli* components: core RNA polymerases and the products of *glnF*, *glnG* and *glnL*. *Proc. Natl. Acad. Sci., USA* 82, 8452-8457.
- Hunter, I.S. (1985). Gene cloning in *Streptomyces*. In: "DNA Cloning." II. Glover, D.M. (ed). IRL Press.

- Hunter, I.S. and Baumberg, S. (1989). Molecular genetics of antibiotic formation. In: "Microbial products: New approaches." The Society for General Microbiology Symposium 44. Baumberg, S., Hunter, I.S. and Rhodes, P.M. (eds). Cambridge University Press. pp. 121-162.
- Huo, L., Martin, K.J. and Schleif, R. (1978). Alternative DNA loops regulate the arabinose operon in *Escherichia coli*. Proc. Natl. Acad. Sci. USA 85, 5444-5448.
- Hutter, R. and Eckhardt, T. (1988). Genetic manipulation. In: "Actinomycetes in Biotechnology." Goodfellow, M., Williams, S.T. and Mordarski, M. (eds). Academic Press, London, pp. 89-184.
- Hutton, J.R. (1977). Renaturation kinetics and thermal stability of DNA in aqueous solutions of formamide and urea. Nuc. Acid Res. 4, 3537-3555.
- Igo, M.M., Ninfa, A.J. and Silhavy, T.J. (1989a). A bacterial environmental sensor that functions as a protein kinase and stimulates transcriptional activation. Genes Dev. 3, 598-605.
- Igo, M.M., Ninfa, A.J., Stock, J.B. and Silhavy, T.J. (1989b). Phosphorylation and dephosphorylation of a bacterial activator by a transmembrane receptor. Genes Dev. 3, 1725-1734.
- Ingram, C., Brawner, M., Youngman, P. and Westpheling, J. (1989). *xylE* functions as an efficient reporter gene in *Streptomyces* spp.: use for the study of *galp1*, a catabolite-controlled promoter. J. Bacteriol. 171, 6617-6624.
- Inouye, M. and Delihias, N. (1988). Small RNAs in the prokaryotes: a growing list of diverse roles. Cell 53, 5-7.
- Irani, M., Orosz, L. and Adhya, S. (1983). A control element within a structural gene: the *gal* operon of *Escherichia coli*. Cell 22, 783-788.
- Iserentant, D. and Fiers, W. (1980). Secondary structure of mRNA and efficiency of translation initiation. Gene 9, 1-12.
- Ishida, I., Obinata, M. and Deguchi, T. (1987). Molecular cloning and nucleotide sequence of cDNA encoding hydroxyindole O-methyltransferase of bovine pineal glands. J. Biol. Chem. 262, 2895-2899.
- Janssen, G., Ward, J.M. and Bibb, M. (1989). Unusual transcriptional and translational features of the aminoglycoside phosphotransferase gene (*aph*) from *Streptomyces fradiae*. Genes and Development 3, 415-429.
- Jinks-Robertson, S. and Nomura, M. (1987). Ribosomes and tRNA. In: *Escherichia coli* and *Salmonella typhimurium*". Neidhardt, F.C., Ingraham, J.L., Brooks Low, K., Magasanik B., Schaechter, M. and Umberger, H.E (eds). American Society for Microbiology. pp. 1358-1385.

- Johnson, R.C. and Simon, M.I. (1985). Host protein requirements for *in vitro* site-specific DNA inversion. *Cell* 46, 531-539.
- Kahmann, R., Rudt, F., Koch, C. and Mertens, G. (1985). G inversion in bacteriophage Mu DNA is stimulated by a site within the invertase gene and a host factor. *Cell* 41, 771-780.
- Kaneko, M., Yamaguchi, A. and Sawai, T. (1985). Energetics of tetracycline efflux system encoded by Tn10 in *Escherichia coli*. *FEBS Lett*, 193, 194-198.
- Kassavetis, G.A. and Geiduschek, E.P. (1984). Defining a bacteriophage T4 late promoter: the T4 gene 55 protein suffices for directing late promoter recognition. *Proc. Nat. Acad. Sci. USA* 81, 5101-5105.
- Katz, L., Staver, M., Tuan, J., Brown, D. and Donadio, S. (1990). Genetics of erythromycin synthesis in *Saccharopolyspora erythraea*. In: "Molecular Biology of Streptomyces." Abstracts of the UCLA Symposia on Molecular and Cellular Biology. Bibb, M.J., Losick, R. and Taylor, D. (eds). Colorado, USA.
- Kendall, K.J. and Cohen, S.N. (1988). Complete nucleotide sequence of the *Streptomyces lividans* plasmid pIJ101 and correlation of the sequence with genetic properties. *J. Bacteriol.* 170, 4634-4651.
- Kenney, T.J. and Moran, C.P., Jr. (1987). Organisation and regulation of an operon that encodes a sporulation-essential factor in *Bacillus subtilis*. *J. Bacteriol.* 169, 3329-3339.
- Kinashi, M., Shimaji, M. and Sakai, A. (1987). Giant linear plasmids in *Streptomyces* which code for antibiotic biosynthesis genes. *Nature*, 238, 454-456.
- Kirby, K.S., Fox-Carter, E. and Guest, M. (1967). Isolation of deoxyribonucleic acid and ribosomal ribonucleic acid from bacteria. *Biochem. J.* 104, 258-262.
- Kirby, R. and Hopwood, D.A. (1977). Genetic determination of methylenomycin synthesis by the SCP1 plasmid of *Streptomyces coelicolor* A3(2). *J. Gen. Microbiol.* 98, 239-252.
- Kominek, L.A. (1971). Biosynthesis of novobiocin by *Streptomyces niveus*. *Antimicrob. Agents Chemother.* 1, 123-134.
- Kondo, S., Yasui, K., Natsume, M., Katayama, M. and Marumo, S. (1988). Isolation, physico-chemical properties and biological activity of pamamycin-607, an aerial mycelium-inducing substance from *Streptomyces alboniger*. *J. Antibiot.* 41, 1196-1204.
- Kramer, H., Amouyal, M., Nordheim, A. and Muller-Hill, B. (1988). DNA supercoiling changes the spacing requirement of two *lac* operators for DNA loop formation with *lac* repressor. *EMBO J.* 7, 547-556.

- Krinke L. and Wulff, D.L. (1987). OOP RNA produced from multicopy plasmids, inhibits *cII* gene expression through a RNase III-dependent mechanism. *Genes Dev.* 1, 1005-1013.
- Lawlor, E.J., Baylis, H.A. and Chater, K.F. (1987). Pleiotropic morphological and antibiotic deficiencies result from mutations in a gene encoding a tRNA-like product in *Streptomyces coelicolor*. *Genes and Development* 1, 1305-1310.
- Lee, J.J., Lee, J.P., Keller, P.J., Cottrel, C.J., Chang, C.J., Zahner, H. and Floss, H.G. (1986). Further studies on the biosynthesis of chlorothricin. *J. Antibiot.* 39, 1123-1134.
- Liao, S.M., Wu, T-H., Chiang, C.H., Susskind, M.M. and McClure, W.R. (1987). Control of gene expression in bacteriophage P22 by a small antisense RNA. I. Characterisation *in vitro* of the *P<sub>sar</sub>* promoter and the *sar* RNA transcript. *Genes Dev.* 1, 197-203.
- Liras, P., Villanueva, J.R. and Martin, J.F. (1977). Sequential expression of macromolecule biosynthesis and candididin formation in *Streptomyces griseus*. *J. Gen. Microbiol.* 102, 269-277.
- Liras, P., Astorias, J.A., Martin, J.F. (1990). Phosphate control sequences involved in transcriptional regulation of antibiotic biosynthesis. *TIBTECH.* 8, 184-189.
- Losick, R. and Pero, J. (1981). Cascades of sigma factors. *Cell* 25, 582-584.
- McCormick, J.R.D. (1957). A new family of antibiotics: the demethyl-tetracyclines. *J. Amer. Chem. Soc.* 79, 4561-4564.
- McCormick, J.R.D. (1960). Biosynthesis of the tetracycline: an integrated biosynthetic scheme. In: "Antibiotics - Advances in Research Production and Clinical Use." Herold, M. and Gabriel, Z. (eds). Czechoslovak Medical Press, Prague, pp. 556-573.
- McCormick, J.R.D. and Jensen, E.R. (1969). Biosynthesis of tetracyclines. XII. Anhydromethylchlorotetracycline from a mutant of *Streptomyces aureofaciens*. *J. American Chem. Soc.* 91, 206-221.
- McLaughlin, J.R., Murray, C.L. and Rabinowitz (1981). Unique features of the ribosome binding site sequence of the Gram-positive *Staphylococcus aureus*  $\beta$ -lactamase gene. *J. Biol. Chem.* 256, 11283-11291.
- Magasanik, B. (1989). Gene regulation from sites near and far. *The New Biologist*, 1, 247-351.
- Magasanik, B. and Neidhardt, F.C. (1987). Regulation of carbon and nitrogen utilisation. In: "*Escherichia coli* and *Salmonella typhimurium*". Neidhardt, F.C., Ingraham, J.L., Brooks Low, K., Magasanik B., Schaechter, M. and Umberger, H.E (eds). Cellular and Molecular Biology. American Society of Microbiology, Washington DC. pp. 1318-1325.

- Majumdar, M.K. and Majumdar, S.K. (1971). Relationship between alkaline phosphatase and neomycin formation in *S. fradiae*. J. Biochem. 122, 397-404.
- Makino, K., Shinagawa H., Amemura, M. and Nakata A. (1986). Nucleotide sequence of the *phoB* gene, the positive regulatory gene for the phosphate receptor of *Escherichia coli* K-12. J. Mol. Biol. 190, 37-44.
- Makino, K., Shinagawa, H., Amemura, M., Kimura, S., Nakada, A. and Ishihawa, A. (1988). Regulation of the phosphate regulon of *Escherichia coli*: activation of *pstS* transcription by PhoB protein *in vitro*. J. Mol. Biol. 203, 85-95.
- Malpartida, F. and Hopwood, D.A. (1984). Molecular cloning of the whole biosynthetic pathway of a *Streptomyces* antibiotic and its expression in a heterologous host. Nature 309, 462-464.
- Malpartida, F. and Hopwood, D.A. (1986). Physical and genetic characterization of the gene cluster for the antibiotic actinorhodin in *Streptomyces coelicolor* A3(2). Mol. Gen. Genet. 205, 66-71.
- Malpartida, F., Hallam, S.E., Kieser, H.M., Motamedi, H., Hutchinson, C.R., Butler, M.J., Sugden, D.A., Warren, M., McKillop, C., Bailey, C.R., Humphreys, G.O. and Hopwood, D.A. (1987). Sequence homology between *Streptomyces* genes coding for synthesis of different polyketides and its use to clone antibiotic biosynthetic genes. Nature 325, 818-820.
- Mandecki, W. and Reznikoff, W.S. (1982). A *lac* promoter with a changed distance between -10 and -35 regions. Nuc. Acid Res. 10, 903-912.
- Maniatis, T., Fritsch, E.F. and Sambrook, J. (1989). In: "Molecular cloning: A laboratory manual." II. Cold Spring Harbor Laboratory, New York.
- Martin, J.F. & Demain, A. (1980). Control of antibiotic biosynthesis. Microbiol. Rev. 44, 230-251.
- Martin, J.F. and Demain A.L. (1977). Effects of exogenous nucleotides on the candicidin fermentation. Can. J. Microbiol. 23, 1334-1339.
- Martin, J.F., Daza, A., Asturias, J.A., Gil, J.A. and Liras, P. (1988). Transcriptional control of antibiotic biosynthesis at phosphate regulated promoters and cloning of a gene involved in the control of the expression of multiple pathways in *Streptomyces*. In: "Biology of Actinomycetes '88." Okami, Y. (ed).
- Maxam, A.M. and Gilbert, W. (1977). A new method for sequencing DNA. Proc. Nat. Acad. Sci. USA 74: 560-564.
- Meinkoth, J. and Wahl, G. (1984). Hybridisation of nucleic acids immobilised on solid supports. Analytical Biochem. 138, 285-290.

- Meinkoth, J. and Wahl, G. (1984). Hybridisation of nucleic acids immobilised on solid supports. *Analytical Biochem.* 138, 285-290.
- Miller, V.L., Taylor, R.K. and Mekalanos, J.J. (1987). Cholera toxin transcriptional activator ToxR is a transmembrane DNA binding protein. *Cell* 48, 271-279.
- Mitrovic, M. and Schildknecht, E.G. (1974). Anticoccidial activity of lasalocid in chicks. *Poultry Science* 53, 1448-1452.
- Miyake, K., Kuzuyama, T., Horinouchi, S. and Beppu, T. (1990). The A factor-binding protein of *Streptomyces griseus* negatively controls streptomycin production and sporulation. *J. Bacteriol.* 172, 3003-3008.
- Mizosawa, S., Nichimora, S. and Seda, F. (1986). Improvement of the dideoxy chain termination method of DNA sequencing by use of deoxy-7-deazaguanosine triphosphate in place of dGTP. *Nuc. Acid Res.* 14, 1319-1324.
- Mizuno, T. and Mizushima, S. (1986). Characterisation by deletion and localised mutagenesis *in vitro* of the promoter region of the *Escherichia coli ompC* gene and importance of upstream DNA domain in positive regulation of the *ompR* protein. *J. Bacteriol.* 168, 86-95.
- Moran Jr., C.P., Lang, N.N., Le Grice, S.F.J., Lee, G., Stephens, M., Sonenshein, A.L., Pero, J. and Losick, R. (1982). Nucleotide sequences that signal the initiation of transcription and translation in *Bacillus subtilis*. *Mol. Gen. Genet.* 186, 339-346.
- Motamedi, H. and Hutchinson, C.R. (1987). Cloning and heterologous expression of a gene cluster for the biosynthesis of tetracenomycin C, the anthracycline antitumour antibiotic of *Streptomyces glaucescens*. *Proc. Nat. Acad. Sci. USA* 84, 4445-4449.
- Motamedi, H., Wendt-Pienkowski, E. and Hutchinson, C.R. (1986). Isolation of tetracenomycin C non-producing *Streptomyces glaucescens* mutants. *J. Bacteriol.* 167, 575-580.
- Mountain, A., McChesney, J., Smith, M.C.M. and Baumberg, S. (1986). Gene sequence encoding early enzymes of arginine synthesis within a cluster in *Bacillus subtilis*, as revealed by cloning in *Escherichia coli*. *J. Bacteriol.* 165, 1026-1028.
- Mountain, A., Smith, M.C.M. and Baumberg, S. (1990). Nucleotide sequence of the *Bacillus subtilis argF* gene encoding ornithine transferase. *Nuc. Acid Res.* 18, 4594.
- Mulligan, M.E., Brosius, E.J. and McClure, W.R. (1985). Characterisation *in vitro* of the effect of spacer length on the activity of *Escherichia coli* RNA polymerase at the *tac* promoter. *J. Biol. Chem.* 260, 3529-3538.



- Nakane, M., Yokoyama, E. and Deguchi, T. (1983). Species heterogeneity of pineal hydroxyindole-O-methyl transferase. *J. Neurochem.* 40, 790-796.
- Nihari, T., Shimuzu, Y., Kim, H.S. and Yamada, Y. (1988). Structure-activity relationship of virginiae butanolide C, an inducer of virginiamycin production in *Streptomyces virginiae*. *J. Antibiot.* 42, 1828-1837.
- Nilsson, L., Vanet, A., Vijgenboom, E. and Bosch, L. (1990). The role of FIS in *trans* activation of stable RNA operons of *E. coli*. *EMBO Journal* 9, 727-734.
- Ninfa, A.J., Ninfa, E.G., Lupas, A., Stock, A., Magasanik R. and Stock, J. (1988). Crosstalk between bacterial chemotaxis signal transduction proteins and the regulators of transcription of the N<sub>2</sub> regulon: evidence that nitrogen assimilation and chemotaxis are controlled by a common phosphotransfer mechanism. *Proc. Natl. Acad. Sci. USA* 85, 5492-5496.
- Ninomija, M., Yamada, Y., Onitsuka, M., Tanaka, Y., Maeda, T. and Maruyama, H.B. (1980). Biochemically active substances from microorganisms. Establishment of several *in vitro* antiinflammatory probes applicable to microbial broths and the effects of non-steroidal antiinflammatory drugs and antibiotics on them. *Chem. Pharm.* 25, 2553-2564.
- Ochi, K. (1986). Occurrence of the stringent response in *Streptomyces* spp. and its significance for the initiation of morphological and physiological differentiation. *J. Gen. Microbiol.* 132, 2621-2631.
- Ochi, K. (1987). Changes in nucleotide pools during sporulation of *Streptomyces griseus* in submerged culture. *J. Gen. Microbiol.* 133, 2787-2795.
- Ochi, K. (1990a). *Streptomyces relC* mutants with an altered ribosomal protein ST-L11 and genetic analysis of a *Streptomyces griseus relC* mutant. *J. Bacteriol.* 172, 4008-4016.
- Ochi, K. (1990b). A relaxed (*rel*) mutant of *Streptomyces coelicolor* A3(2) with a missing ribosomal protein lacks the ability to accumulate ppGpp, A-factor and prodigiosin. *J. Gen. Microbiol.* 136, 2405-2412.
- Ogawa, H., Gomi, T., Mueckler, M.M., Fujioka, M., Backlund, P.S., Aksamit, R.P., Unson, C.G. and Cantoni, G.L. (1987). Amino acid sequence of S-adenosyl-L-homocysteine hydrolase from rat liver as derived from the cDNA sequence. *Proc. Natl. Acad. Sci. USA* 84, 719-723.
- Ohnuki, T., Katoh, T., Imanka, T. and Aiba, S. (1985). Molecular cloning of tetracycline resistance genes from *Streptomyces rimosus* in *Streptomyces griseus* and characterization of the cloned genes. *J. Bacteriol.* 161, 1010-1016.

- Omura, S. (1986). Philosophy of new drug discovery. *Microbiol. Rev.* 50, 259-279.
- Pero, J., Lee, G., Moran, C.P., Lang, N. and Losick, R. (1982). Promoters controlled by novel sigma factors in *Bacillus subtilis*. In: "Promoters, structure and function." Rodriguez, R.L. and Chamberlin, M.J. (eds). Praeger Publishers. pp. 283-293.
- Petersen, G.B., Stockwell, P.A. and Hill, D.F. (1988). Messenger RNA recognition in *Escherichia coli*: a possible second site of interaction with 16S ribosomal RNA. *EMBO J.* 7, 3957-3962.
- Pigac, J. and Alacevic, M. (1979). Mapping of oxytetracycline genes in *Streptomyces rimosus*. *Period Biol.* 81, 575-582.
- Piret, J.M. and Chater, K.F. (1985). Phage-mediated cloning of *bldA*, a region involved in *Streptomyces coelicolor* morphological development and its analysis by genetic complementation. *J. Bacteriol.* 163, 965-972.
- Platt, T. and Bear, D.G. (1983). Role of RNA polymerase, rho factor and ribosomes in transcription termination. In: "Gene Function in Prokaryotes". Beckwith, J., Davis, J.A. and Gallant, J.A. (eds). New York. Cold Spring Harbor Laboratory. pp. 123-161.
- Ptashne, M. (1986). "A Genetic Switch: Gene Control and Phage Lambda." Blackwell Scientific Publications.
- Rebollo, A., Gil, J.A., Asturias, J.A., Liras, P. and Martin J.F. (1989). Cloning and characterisation of a phosphate regulated promoter involved in phosphate control of candicidin biosynthesis. *Gene* 79, 47-58.
- Reynes, J-P., Calmeb, T., Procourt, D. and Tiraby, G. (1988). Cloning expression in *Escherichia coli* and nucleotide sequence of a tetracycline resistance gene from *Streptomyces rimosus*. *J. Gen. Microbiol.* 134, 585-598.
- Reznikoff, W.S., Winter, R.B. and Hurley, C.K. (1987). The location of the repressor binding sites in the *lac* operon. *Proc. Natl. Acad. Sci. USA* 71, 2314-2318.
- Rhodes, P.M., Winskill, N., Friend, E.J. and Warren, M. (1981). Biochemical and genetic characterization of *Streptomyces rimosus* mutants impaired in oxytetracycline biosynthesis. *J. Gen. Microbiol.* 124, 329-338.
- Rhodes, P. M., Hunter, I. S., Friend, E. J. and Warren, M. (1984). Recombinant DNA methods for the oxytetracycline producing *S. rimosus*. *Biochem. Soc. Transac.* 12, 586-587.
- Richardson, S.M.H., Higgen, C.F. and Lilley, D.M. (1980). DNA supercoiling and the *leu-500* promoter mutation of *Salmonella typhimurium*. *EMBO J.* 7, 1863-1869.

- Roberts, M. and Rabinowitz, J.C. (1988). The role of ribosomal protein S1 in species-specific protein synthesis. In: "Genetics and Biotechnology of Bacilli." II. Ganesan, A.T. and Hoch, J.A. (eds). Academic Press, California. pp 101-108.
- Roberts, M.W. and Rabinowitz, J.C. (1989). The effect of *Escherichia coli* ribosomal protein S1 on the translational specificity of bacterial ribosomes. *J. Biol. Chem.* 266, 2228-2235.
- Ronson, C.W., Nixon, B.T. and Ausubel, F.M. (1987). Conserved domains in bacterial regulatory proteins that respond to environmental stimuli. *Cell* 49, 579-581.
- Rose, M. and Botstein, D. (1983). Structure and function of the yeast *URAB* gene: differentially regulated expression of hybrid  $\beta$ -galactosidase from overlapping coding sequences in yeast. *J. Mol. Biol.* 170, 883-904.
- Rudd, B.A.M. and Hopwood, D.A. (1979). Genetics of actinorhodin biosynthesis by *Streptomyces coelicolor* A3(2). *J. Gen. Microbiol.* 114, 35-43.
- Sanger, F., Nicklen, S. and Coulson, A.R. (1977). DNA sequencing with chain terminating inhibitors. *Proc. Nat. Acad. Sci.* 74, 5463-5467.
- Sankaran, L. and Pogell, B.M. (1975). Biosynthesis of Puromycin *Streptomyces alboniger*: regulation and properties of O-methyl puromycin O-methyl transferase. *Antimicrob. Agents Chemother.* 8, 721-732.
- Sasse-Dwight, S. and Gralla, J.D. (1988). Probing the *Escherichia coli* *glnALG* upstream activation mechanism *in vivo*. *Proc. Natl. Acad. Sci. USA* 85, 8934-8938.
- Schumard, R.F. and Callendar, M.E. (1968). Monensin, a new biologically active compounds. VI. Anticoccidial activity. *Antimicro. Agent. Chemo.* 1967, 369-372.
- Seno, E.T. and Baltz, R.H. (1988). Structural organization and regulation of antibiotic biosynthesis and resistance genes in actinomycetes. In: "Regulation of Secondary Metabolism in Actinomycetes". Shapiro, S. (ed.). CRC Press. pp 1-48.
- Sherman, D.H., Malpartida, F., Bibb, M.J., Kieser, H.M., Bibb, M.J. and Hopwood, D.A. (1989). Structure and deduced function of the granaticin-producing polyketide synthase gene cluster of *S. violaceoruber* Tu22.
- Shine, J. and Dalgarno, L. (1974). The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc. Nat. Acad. Sci.* 71, 1342-1346.

- Shine, J. and Delgarno, L. (1975). Determinant of cistron specificity in bacterial ribosomes. *Nature* 254, 34-35.
- Simons, R.W. and Kleckner, N. (1983). Translational control of IS10 transposition. *Cell* 34, 683-689.
- Simpson, R.B. (1982). Evidence for close contacts between RNA polymerase and promoter DNA. In: "Promoters Structure and Function." Rodriguez, R.L. and Chamberlin, M.J. (eds). Praeger Publishers. pp. 164-183.
- Sinclair, A.C., Schenk, J.R., Post, G.G., Cardinal, E.V., Burakas, S. and Fricke, H.H. (1962). Chelocaldin, a new broad-spectrum antibiotic. II. Isolation and characterisation. *Antimicro. Agents and Chemo.* 1, 592-595.
- Sinclair, R.B. and Bibb, M.J. (1988). The repressor gene (*c*) of the *Streptomyces* phage  $\phi$ C31: Nucleotide sequence, analysis and functional cloning. *Mol. Gen. Genet.* 213, 269-277.
- Smith, I. (1989). The initiation of sporulation. In: "Regulation of prokaryotic development". Smith I., Slepecky R. and Setlow P. (ed). American Society for Microbiology, Washington D.C, pp. 185-210.
- Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98, 503-517.
- Stanzak, R., Matsushima, P., Baltz, R.H. and Rao, R.N. (1986). Cloning and expression in *Streptomyces lividans* of clustered erythromycin biosynthesis genes from *Streptomyces erythreus*. *Biotechnology* 4, 229-232.
- Stefano, J.E. and Gralla, J.D. (1982). Spacer mutations in the *lacpS* promoter. *Proc. Nat. Acad. Sci., USA*, 79, 1069-1072.
- Stock, J.B., Ninfa, A.J. and Stock, A.M. (1989). Protein phosphorylation and regulation of adaptive responses in bacteria. *Microbiological Reviews* 53, 450-490.
- Stormo, G.D. (1986). Translation initiation. In: "Maximising gene expression". Reznikoff, W.S. and Gold, L. (ed), pp. 195-224.
- Stormo, G.D., Schneider, T.D. and Gold, L.M. (1982). Characterization of translation initiation sites in *E. coli*. *Nuc. Acid Res.* 10, 2971-2996.
- Strauch, E., Takano E., and Bibb, M.J. (1990). The stringent response in *Streptomyces coelicolor* A3(2). *Molecular Microbiology* (In press).
- Suzuki, Y. and Yamada, T. (1988). The nucleotide sequence of 16S rRNA from *Streptomyces lividans* TK21. *Nuc. Acid Res.* 16, 370.

- Takahashi, H., Tanaka, K. and Shiina, T. (1988). Genetic constituents of *rpoD* gene homologues in *Streptomyces* strains. In: "Seventh International Symposium on Biology of Actinomycetes." Okami, Y., Beppu, T. and Ogarwara H. (eds). Scientific Societies Press, Tokyo.
- Tanaka, K., Shiina, T. and Takahashi, H. (1988). Multiple principle sigma factor homologs in eubacteria: identification of the *rpoD* box. *Science* 242, 1040-1042.
- Thomas, R. and Williams, D.J. (1983). Oxytetracycline biosynthesis: mode of incorporation of [1-<sup>13</sup>C] and [1,2-<sup>13</sup>C<sub>2</sub>] acetate. *J. Chem. Soc., Chem. Commun.* 12, 677-679.
- Tohyama, H., Okami, Y. and Umezawa, H. (1986). Negative control for the expression of streptomycin resistance gene from streptomycin-producing *Streptomyces griseus*. *J. Antibiot.* 39, 1505-1507.
- Torriani, A. and Rothman, F. (1961). Mutants of *Escherichia coli* constitutive for alkaline phosphatase. *J. Bacteriol.* 81, 835-836.
- Turner, W.B. and Aldridge, D.C. (1983). "Fungal Metabolites." II. Academic Press: London and New York.
- Vancurova, I., Volc, J., Flieger, M., Neuzil, J., Novotna, J., Vlach, J. and Behal, V. (1988). Isolation of pure anhydrotetracycline oxygenase from *Streptomyces aureofaciens*. *Biochem. J.* 253, 263-267.
- Ward, J.M., Janssen, G.R., Kieser, T., Bibb, M.J., Buttner, M.J. and Bibb, M.J. (1986). Construction and characterisation of a series of multicopy promoter probe plasmid vectors for *Streptomyces* using aminoglycoside phosphotransferase gene from Tn5 as indicator. *Mol. Gen. Genet.* 203, 468-478.
- Westpheling, J. and Brawner, M. (1989). Two transcribing activities are involved in expression of *Streptomyces galactose* operon. *J. Bact.* 171, 1355-1361.
- Westpheling, J.E., Raines, M. and Losick, R. (1985). RNA polymerase heterogeneity in *Streptomyces coelicolor*. *Nature* 313, 22-27.
- Wierenga, R.K. and Hol, W.G.J. (1983). Predicted nucleotide-binding properties of p21 protein and its cancer-associated variant. *Nature* 302, 842-844.
- Wildermuth, H. (1970). Development and organization of the aerial mycelium in *Streptomyces coelicolor*. *J. Gen. Microbiol.* 60, 43-50.
- Yamada, M., Makino, K., Amemura, M., Shinagawa, H. and Nakara, A. (1989). Regulation of the phosphate regulon of *Escherichia coli* Analysis of mutant *phoB* and *phoR* genes causing different phenotypes. *J. Bacteriol.* 171, 5601-5605.

- Yanisch-Perron, C., Vieira, J. and Messing, J. (1985). Improved M13 phage cloning vectors and host strains nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* 33, 103-119.
- Yoshida, K., Okamoto, M., Umehara, K., Iwami, M., Kosaka, M., Aoki, H. and Imanuka, H. (1982). Studies on new vasodilators WS-1128A and WS-1228-B1. Discovery, taxonomy, isolation and characterisation. *J. Antibiot.* 35, 151-156.
- Yue, S., Motamedi, H., Wendt-Pienkowski, E. and Hutchinson, C.R. (1986). Anthracycline metabolites of tetracenomycin C-nonproducing *Streptomyces glaucescens* mutants. *J. Bacteriol.* 167, 581-586.
- Zalkin, H. and Ebbole, D.J. (1988). Organisation and regulation of genes encoding biosynthetic enzymes in *Bacillus subtilis*. *J. Biol. Chem.* 263, 1595-1598.
- Zengel, J.M., Archer, R.H. and Lindahl, L. (1984). The nucleotide sequence of the *Escherichia coli* *fus* gene, coding for elongation factor G. *Nuc. Acid Res.* 12, 2181-2192.
- Zhang, H-I, He, X-g, Adefarati, A., Galluci, J., Cole, S.P., Beale, J.M., Keller, P.J., Chang, C-j. and Floss, H.G. (1990). Mutactin, a novel polyketide from *Streptomyces coelicolor*. Structure and biosynthetic relationship to actinorhodin. *J. Organ. Chem.* 55, 1682-1684.
- Zukowsa, M.M., Gaffney, D.F., Speck, D., Kauffmann, M., Findeli, A., Wisecup, A. and Lecocq, J.P. (1983). Chromogenic identification of genetic regulatory signals in *Bacillus subtilis* based on expression of a cloned *Pseudomonas* gene. *Proc. Nat. Acad. Sci. USA* 80, 1101-1105.